The Perception of Contrastive Stress in Vocoded Speech:

Implications for Cochlear Implant Users

Alix Klang

Submitted under the supervision of Dr. Arlene Carney to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Arts, *summa cum laude* in Speech-Language-Hearing Sciences.

May 13, 2017

Acknowledgements

Thesis Supervisor:

Dr. Arlene Carney

Readers:

Dr. Evelyn Davies-Venn

Dr. Peter Watson

Abstract

Vocoded speech is used frequently to simulate cochlear implant processing in normal-hearing listeners (Loebach, 2007; Oxenham & Kreft, 2014). Speech signals are band-pass filtered at the first stage of processing and outputs of each of these filters are subsequently low-pass filtered. Varying the values of low-pass filter cutoffs influences periodicity cues conveyed through vocoded speech, with higher values passing more periodicity cues and lower levels reducing these cues. We examined the effect that limiting periodicity cues have in the perception of contrastive stress. Previous research showed that both spectral and temporal features of sentence stimuli are varied when speakers use contrastive stress (Cooper, Eady, & Mueller, 1985). We were primarily interested in how variations in spectral, temporal, and intensity features of vocoded speech influenced the perception of contrastive stress. We presented stimuli to participants in four conditions of a contrastive stress experimental test: unprocessed, natural speech and three processed speech conditions with low-pass filter cutoffs at 50 Hz, 160 Hz, and 250 Hz. Participants were assigned to only one filter condition. We also evaluated how talker and syntactic place influenced the perception of contrastive stress. There were four talkers and four syntactic stress conditions. All participants received were sentence stimuli for each talker and syntactic stress condition. We found significant perception effects for filter condition, talker, and place, as well as a significant interaction of talker by place. These outcomes suggest that the perception of contrastive stress in vocoded speech is influenced by a number of factors, including: (1) the availability of periodicity cues primarily conveyed by fundamental frequency, (2) talker variation, and (3) syntactic place of the stressed word in a sentence. The results may partially explain the range of performance experienced by cochlear implant users in tasks that test the perception of suprasegmental cues.

Table of Contents

List of Tables

List of Figures

List of Appendices

I. Introduction

a. Cochlear Implants

Cochlear implants (CIs) produce the sensation of sound in individuals with severe-to-profound hearing loss due to cochlear damage by providing electrical stimulation to the auditory nerve (Hughes, 2013a; Loebach, 2007; Qin & Oxenham, 2005). The primary components of a cochlear implant are the microphone, speech processor, external transmitter, implanted receiver/stimulator, and electrode array. First, the microphone picks up sound in the environment, such as speech, and directs it to the speech processor. The speech processor transforms the speech signal through band-pass filtering and rectification. This output is further directed to the external transmitter, a high-bandwidth radio-frequency transmitter, which sends the signal to a surgically implanted receiver/stimulator in the temporal bone. The receiver/stimulator passes the signal to an electrode array, inserted within the cochlea. The electrode array delivers the signal, in the form of electrical pulses, to stimulate the auditory nerve (Dorman & Wilson, 2004).

The speech processor in modern cochlear implants frequently implements the signal processing strategy of CIS, or continuous interleaved sampling (Wilson, 2015; Wilson, Dorman, Gifford, & McAlpine, 2016; Qin & Oxenham, 2005). CIS uses multiple channels of speech processing and the output of each channel is directed to a corresponding site of stimulation, a specific electrode on the array. Each channel includes a band-pass filter, energy detector, nonlinear mapping function, and multiplier (Wilson, 2015; Wilson et al., 2016). Available cochlear implants today contain between 6 and 24 channels (Loebach, 2007). The frequency response of each band-pass filter of a speech processor differs, ranging from low to high frequencies along a logarithmic scale. For example, with a six-channel processor, the pass bands of each filter may be 300-494 Hz, 494-814 Hz, 814-1342 Hz, 1342-2210 Hz, 2210-3642 Hz, and 3642-6000 Hz (Wilson, 2015; Wilson et al., 2016). The spacing of filters in a speech processor, with any number of channels, simulates the frequency map of the cochlea from base to apex, with high frequencies at the base and low frequencies at the apex (Dorman & Wilson, 2004; Wilson, 2015; Xu & Pfingst, 2008). This tonotopic organization of the cochlea occurs due to the graded mechanical properties of the basilar membrane within the cochlea, which stimulate hair cells and their corresponding neurons in response to specific frequencies. High frequencies are analyzed by the narrow and stiff base of the basilar membrane and low frequencies are analyzed by the wide and flexible apex. This tonotopic organization is maintained from the cochlea along the auditory pathway to the temporal lobe of the cortex (Dorman &Wilson, 2004; Wilson, 2015; Wilson et al., 2016).

In a cochlear implant that uses a CIS processing strategy, the energy detector, or envelope detector, follows the band-pass filter and consists of both a rectifier and low-pass filter. The energy detector typically uses either a full-wave or half-wave rectifier. A full-wave rectifier converts a sound signal into electrical pulses using both the positive and negative half cycles of the signal, whereas a half-wave rectifier uses only the positive half cycles. Following the rectifier is the low-pass filter. The low-pass filter sets the cutoff frequency of the envelope detector. This frequency cutoff is typically between 200 and 400 Hz, with 400 Hz being the most common (Wilson, 2015; Wilson et al., 2016).

Following the energy detector is the nonlinear mapping function. Similar to the band-pass filter, the nonlinear mapping function is logarithmic in nature and is used within each channel to compress the dynamic range of sounds. The wide dynamic range of environmental sounds, from 90 to 100 dB, is compressed into a narrow electrical dynamic range, between 5 and 20 dB. This narrow dynamic range depends on both the patient and the number of electrodes within their implant. The purpose of the nonlinear mapping function is to allow for CI users to perceive low-level sounds as soft and high-level sounds as comfortably loud. Further, this output is employed by the multiplier to modulate a train of electrical pulses to each electrode (Wilson, 2015; Wilson et al., 2016).

In the next stage, the multiplier modulates the train of balanced biphasic pulses (Faulkner, Rosen, & Smith, 2000; Wilson, 2015; Wilson et al., 2016; Qin & Oxenham, 2005; Xu & Pfingst, 2008). These biphasic pulses contain both negative and positive components. In order to mitigate crosstalk among electrodes, there can be no simultaneous or overlapping stimulation of electrodes. This is referred to as interleaved stimulation of electrodes, which gives rise to the name continuous interleaved sampling. Without interleaved stimulation, the representation of place cues within a cochlear implant would be reduced due to perceptual distortions. Further, the pulse rate is required to be at least four times the highest frequency of the modulation waveform. The average cutoff frequency, as stated above, is 400 Hz. As a result, the pulse rate for each channel and subsequent electrode will be approximately 1600 pulses per second. This serves to provide an undistorted representation of the signal waveform in the response of the auditory nerve (Dorman & Wilson, 2004; Wilson, 2015; Wilson et al., 2016).

The electrode array that is inserted within the scala tympani of the cochlea provides electrical current to the basilar membrane to directly depolarize auditory neurons in the spiral ganglion. This serves to compensate for inner hair cell loss by replacing neurotransmitters with electrical current to generate neural action potentials. Multiple sites of stimulation allow for encoding sound frequencies. Low-center-frequency and high-center-frequency channels are directed to different sites along the electrode array. Low-frequency signals stimulate more apical electrodes and high-frequency signals stimulate more basal electrodes. In addition to the place code, sound frequency is encoded through temporal code. In an undamaged cochlea, neurons phase-lock to the period of an acoustic wave. With cochlear implants, the stimulation of each electrode at different rates produces different pitches, or perceptual interpretations of frequency (Dorman & Wilson, 2004; Hughes, 2013a; Hughes, 2013b; Wilson, 2015).

b. Vocoded Speech: Cochlear Implant Simulation

Vocoded speech is a form of synthesized speech meant to simulate the output of cochlear implants (Loebach, 2007). In many research projects, vocoded speech is used to test normal hearing listeners with signals that have been modulated in a manner similar to that of a cochlear implant speech processor. The three primary components of vocoded speech include frequency channel, amplitude envelope, and carrier signal. First, with frequency channels, a series of band-pass filters are used to divide an acoustic signal into different frequency bands. As with cochlear implants, the band-pass filters used within simulations are broad in order to comprise the range of the frequency spectrum. Each band-pass filter contains a specific energy profile, or spectrum of frequencies. The values of each band-pass filter should approximate the values used within a cochlear implant sound processor. Further, the number of bands may vary from simulation to simulation, depending on the cochlear implant being modeled (Loebach, 2007; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Qin & Oxenham, 2005).

Once the acoustic signal has passed through the series of band-pass filters, the amplitude envelope must be derived from each band (Loebach, 2007; Shannon et al., 1995). Similar to the cochlear implant sound processor, the envelope is extracted through half-wave rectification and then low-pass filtered (Friesen, Shannon, Baskent, & Wang, 2001; Shannon et al., 1995;

Whitmal, Poissant, Freyman, & Helfer, 2007). The low-pass filter cutoff frequency is typically set at 400 Hz or below (Gonzales & Oliver, 2005; Qin and Oxenham, 2005). Higher frequency cutoffs are better able to preserve smaller changes in temporal fluctuations. As a result, low-pass filters with high cutoff frequencies are able to maintain basic pitch information, if the pitch falls within the pass band (Loebach, 2007).

Once the signal has passed through the low-pass filter, it is subsequently replaced with a modulated synthetic source, known as the carrier signal (Loebach, 2007). This carrier signal may either be white noise or sinusoid waves (Dorman, Loizou, & Rainey, 1997; Gonzales & Oliver, 2005; Loeback, 2007; Shannon et al., 1995; Souza & Rosen, 2009; Whitmal et al., 2007). Within each channel, the modulated signal will be band limited with the corresponding band-pass filter (Shannon et al., 1995). By replacing the spectral content of each band with a different signal, residual spectral information is removed. This serves to make an effective cochlear implant simulation (Loebach, 2007).

c. Noise Carriers versus Sine Carriers

Noise carriers are randomly varying signals with rapidly fluctuation envelopes (Gonzales & Oliver, 2005; Souza & Rosen, 2009). As a result, noise vocoders do not provide extensive information on periodic structure. This serves to remove a majority of spectral detail from each channel (Gonzales & Oliver, 2005; Laneau, Moonen, & Wouters, 2005; Loebach, 2007; Shannon et al., 1995). On the other hand, with modulated noise carriers, temporal and amplitude cues remain within the spectral band (Shannon et al., 1995). When the envelope filter cutoff for a noise carrier is greater than the fundamental frequency of a talker, periodicity and intonation cues are signaled through amplitude modulations (Souza & Rosen, 2009; Qin & Oxenham, 2005).

However, due to the inherent amplitude fluctuations of the noise signal, the amplitude modulations may not be very deep and lead to percepts of weak pitch (Souza & Rosen, 2009). Subsequently, noise carriers may over-represent the information of each band. This results in a simulation of broad electrodes with contiguous channels. Unfortunately, the electrodes of actual cochlear implants tend to be narrow and less contiguous, with roll off on both sides due to current spread and electrical diffusion (Loebach, 2007).

With a modulated sine carrier, periodic structure is maintained throughout substantial portions of the signal. The spectrum of each channel output contains a series of harmonic-like spectral components centered at the carrier frequency (Gonzales & Oliver, 2005; Souza & Rosen, 2009). Because the sine carrier is focused at each channel center frequency, intensity rolls off on either side for the harmonic-like spectral components (Loebach, 2007). These sidebands provide additional detection cues by imposing a periodic temporal structure and by producing percepts of strong pitch (Whitmal et al., 2007). Intelligibility may increase because the sine carrier provides cues of periodicity and intonation to a listener. Further, with sine carriers, low modulation rates with periodic information improve intelligibility. On the other hand, aperiodic sounds, which contain higher-rate fluctuations, cue the presence of aperiodic energy. Consequently, these fluctuation rates depend on the filter cutoff of the envelope. When the envelope cutoff for a sine carrier is greater than the talker's fundamental frequency, the envelopes will contain amplitude fluctuations that correspond in rate to the voice pitch of the talker for periodic sounds. When the envelope cutoff is lower than the talker's fundamental frequency, speech periodicity will be removed due to the allowance of only slower rate fluctuations (Souza & Rosen, 2009).

There are other differences between noise-vocoded and sine-vocoded speech: (1) Voice pitch is more prominent in sine-vocoded speech at higher frequencies than in noise-vocoded speech; (2) bigger effects of envelope bandwidth are found for sine-vocoded speech; (3) for sine-vocoded speech, the bandwidth of extracted envelopes is an important determinant of temporal and spectral characteristics, whereas for noise-vocoded speech, the bandwidth of extracted envelopes is an important determinant of temporal characteristics only; and, (4) sine carriers have constant amplitude fluctuations, whereas noise carriers have high envelope fluctuations. The lack of intrinsic envelope fluctuations in sine vocoders facilitates better modulation detection. In contrast, these fluctuations within noise carriers act as interferers and negate the benefits of higher-rate amplitude modulations in speech perception. Subsequently, listeners may experience trouble distinguishing the signal from a randomly occurring component of the carrier. Detection of amplitude modulations and discrimination of modulation rates are made worse with increases in the modulation rates of a noise carrier. Overall differences in performance using sine and noise carriers are attributed in part to the intrinsic fluctuations of noise carriers, which degrade temporal cues. With the limited spectral resolution provided by vocoded speech, listeners are made to rely more on temporal cues. Subsequently, tone carriers appear to be better at reproducing speech envelope fluctuations (Souza & Rosen, 2009; Whitmal et al., 2007).

With tone carriers, higher envelope cutoff frequencies lead to sensitivity to voice pitch variations by normal hearing listeners. Unfortunately, this surpasses the performance of traditional cochlear implant users (Blamey et al., 1984; Souza & Rosen, 2009). Interestingly though, lower cutoff frequencies with sine carriers result in poorer sensitivity to voice pitch in normal hearing listeners. Conversely, high-envelope cutoff frequencies with noise carriers lead to an appropriate sensitivity to voice pitch in normal hearing listeners, when compared with performance of traditional cochlear implant users. Fluctuations from the noise itself are not a factor within real cochlear implants (Souza & Rosen, 2009).

Similar to cochlear implants, high levels of speech understanding are obtained using signal processors with a small number of channels, whether the carrier signal is noise or sine. There is no statistically significant difference in performance beyond eight channels. More channels do not add massively to intelligibility in quiet, while fewer channels would take away from spectral detail (Dorman et al., 1997).

Sine carriers with higher frequency cutoffs exceed the fundamental frequencies of most speakers. This allows for the identification of frequency glides; performance is higher with sine carriers than with vocoders using envelope-modulated noise carriers. Consonant voicing information is better perceived through sine carriers, due to the preservation of periodicity. However, consonant and vowel identification as well as sentence intelligibility are similar through all processors (Faulkner et al., 2000). Modulated sine carriers may be more appropriate due to their metallic-like sound output, which correlates with reports from cochlear implant users that the world sounds more like metallic beep tones than bands of noise (Dorman et al., 1997; Loebach, 2007). Overall, neither sine nor noise vocoders perfectly simulate cochlear implant output.

d. Prosody and Contrastive Stress

A great deal of prosodic information in speech is presented via the fundamental frequency contour (Chatterjee & Peng, 2007; McRoberts, Studdert-Kennedy, & Shankweiler, 1995). This contour is often referred to as speech intonation and is believed to mark linguist contrasts in speech. Temporal features of speech also play an important role in the perception of linguistic contrasts. These temporal features include: temporal envelope, periodicity, and fine temporal structure. Temporal envelope comprises the very slow fluctuations in amplitude of the signal, up to 50 Hz. Periodicity cues can include both periodic and aperiodic variation between 50 and 500 Hz. Fine temporal structure cues include variations in individual wave shapes, observed within single periods of periodic sounds or over short time intervals of aperiodic sounds. Together, marked changes in these temporal features may suggest linguist contrasts within a speech signal (Rosen, 1992).

Contrastive stress, a form of linguistic contrast, is characterized by one or more words within an utterance carrying added stress to denote contrastive emphasis. Cooper et al. (1985) found that both spectral and temporal features of sentence stimuli are influenced by contrastive stress. These investigators asked participants to produce sentences through a question and answer format. A number of the questions were used to elicit contrastive stress on different key words. An example sentence stimulus they provide is "*Chuck* liked the *present* that *Shirley* sent to her *sister*" (Italicized words mark the key words of the sentence). Questions used to elicit contrastive stress on different key words within that sentence include: (1) "Did William or Chuck like the present that Shirley sent to her sister?", (2) "Did Chuck like the letter or the present that Shirley sent to her sister?", (3) Did Chuck like the present that Melanie sent to her sister or the one that Shirley sent?", and, (4) Did Chuck like the present that Shirley sent to her sister or the one she sent to her brother?" Each question is meant to elicit contrastive stress on one of the four key words of the sentence. Thus, question 1 would elicit stress on "Chuck", question 2 would elicit stress on "present", question 3 would elicit stress on "Shirley", and question 4 would elicit stress on "sister". The chosen words were all the subjects or objects in main or relative clauses.

Cooper et al. (1985) discussed the effects of producing words with contrastive stress on both duration and fundamental frequency. The duration of key words was significantly greater when they were the focus of a sentence. Further, duration was influenced by the position of the focus words within the sentences. Focus words that were at the end of the sentence had shorter durations than sentence-initial and sentence-medial focus words. Focus words had no durational effects on other words within the same sentence.

Contrastive stress also had an effect on the fundamental frequency pattern of a sentence. Fundamental frequency increased for each key-word position relative to other words in the sentence, except the first position, in the contrastive stress condition. Similarly, there was little difference in fundamental frequency for the first word position between emphasized and non-emphasized conditions. Cooper et al. (1985) suggested that this was due to the already initial high fundamental frequency for all sentences. Regardless of the contrastive stress condition or neutral stress condition, fundamental frequency decreased throughout the production of each utterance. The greatest drop in fundamental frequency occurred between the first and second key words, regardless of focus location. Lastly, in place of fundamental frequency increases on the focus words, there was post-focus word fundamental frequency decreases (Cooper et al., 1985).

Similar to Cooper et al. (1985), Fry (1958) emphasizes fundamental frequency and duration as acoustical features of stress. Unlike Cooper et al., Fry also noted the role of intensity in signaling stress. Fry had reported that the duration of a stressed word is longer than that on an unstressed word; Cooper et al. had the same outcome. Their arguments diverge on fundamental frequency. Fry suggested that a higher fundamental frequency is present in stressed words. Cooper et al. found that fundamental frequency does not increase on stressed words. Cooper et al. found that fundamental frequency did not increase on stressed words, but rather decreased on following word. Further, Fry indicated that greater intensity signals stress. Accordingly, Fry argues that a higher fundamental frequency, longer duration, and greater intensity will be indicative of stress.

e. Thesis Statement

The purpose of this research was to investigate the perception of contrastive stress in vocoded speech and its implications for cochlear implant users. Thus far, most research on speech perception for individuals with cochlear implants has focused on word and sentence recognition, which depends on the perception of phonemes, with less focus on suprasegmental cues. Those investigators who did study perception of non-segmental cues with CI listeners, as well as with normal hearing (NH) listeners through vocoded speech, addressed the recognition of rising and falling intonation and gender primarily (Chatterjee et al., 2015; Gonzales & Oliver, 2005; Meister, Landwehr, Pyschny, Walger, & von Wedel, 2009; Most & Peled, 2007; Peng, Chatterjee, & Lu, 2012; Schvartz & Chatterjee, 2011). Contrastive stress has only been studied minimally, limiting knowledge on how or if individuals with cochlear implants recognize emphasis on new and/or important information in sentences (Meister et al., 2011).

We wanted to simulate the range of periodicity cues that CI listeners might receive through their devices. We varied the values of the low-pass filter cutoff at the output of 16-band-pass filters in our simulation to determine the impact of limiting fundamental frequency information for the perception of contrastive stress. This limitation of periodicity cues would not affect durational and intensity cues for contrastive stress as directly. As low-pass cutoff filter frequency increases, the availability of periodicity cues will increase as well. We are specifically investigating the periodicity cue to see how participants receive voice pitch. If fundamental frequency is a primary cue for contrastive stress, we hypothesize that increasing the frequency of the low-pass filters cutoff in vocoded speech will be associated with increased performance by NH participants on a contrastive stress test. If fundamental frequency is not an important cue for the perception of contrastive stress, we expect to see changes in performance for changes in duration and intensity regardless of low-pass filter cutoff frequency.

Secondarily, we wanted to determine the specific impact of using multiple talkers with varying fundamental frequencies, rates, and prosodic characteristics on the perception of contrastive stress when we systematically varied the words in sentences that received emphatic stress.

II. Methods

a. Stimuli

Stimuli consisted of 72 simple declarative sentences with the same syntactic frame – 'person', 'verb', and 'object'. Example sentences include "Peg pushed the cart" and "Bob baked the bread." A complete list of the 72 sentences is shown in Appendix A. All 'person' and 'object' words were monosyllabic, whereas the 'verb' words were either monosyllabic or disyllabic. Each word began with a consonant; all 'person' words began with a stop consonant (/p/, /b/, /t/, or /k/). The experimental sentences were designed to be produced twice – once without stress on any single syntactic unit (the non-stressed condition) and once with stress on one of the syntactic units – 'person', 'verb', or 'object'. There were 144 sentences in all.

b. Talkers

Four individual talkers recorded each of the 144 experimental sentences. There were two female talkers (F1 and F2) and two male talkers (M1 and M2). Two talkers (F1 and M1) had a mean age of 24.7 years (F1 and M1) and two talkers had a mean age of 68.8 years (F2 and M2). All talkers were native speakers of General American English.

c. Elicitation and Recording of Stimuli

Stimuli were recorded in a double-walled sound treated room using Praat software (Boersma & Weenink, 2017) with the internal microphone of a MacBook Air. An experimenter sat in the room with the talker and elicited the 144 sentences in a question-answer format. In the unstressed condition, talkers used normal prosody to answer the question "What happened?". In the stressed condition, the experimenter elicited the responses with a series of different questions that would prompt the talker to stress either ‘person’, ‘verb’, or ‘object’. For each sentence stress condition, 24 sentences were recorded (Appendix A).

To elicit sentence stress on the ‘person’ word, the talker was prompted with the question “*Who* [verb] the [object]?” (e.g. “*Who* pushed the cart?”). The talker’s response was then “[*Person*] [verb] the [object]”, with emphasis on the ‘person’ word (e.g. “*Peg* pushed the cart”). To elicit sentence stress on the ‘verb’ word, the talker was prompted with the question “[Person] *did what* to the [object]?” (e.g. “Bob *did what* to the bread?”). The talker’s response was then “[Person] [*verb*] the [object]”, with emphasis on the ‘verb’ word (e.g. “Bob *baked* the bread”). Lastly, to elicit sentence stress on the ‘object’ word, the talker was prompted with the question “[Person] [verb] *what*?” (e.g. “Kate tied *what*?”). The talker’s response was then “[Person] [verb] the [*object*]”, with emphasis on the ‘object’ word (e.g. “Kate tied the *shoes”*).

d. Acoustic Analysis of Stimuli

Two experimenters analyzed the 576 sentences (144 sentences x 4 talkers, 288 stimuli per experimenter) acoustically with Praat software (Boersma & Weenink, 2017). Reliability checks were conducted on 20 randomly selected sentences for each talker. Together, the experimenters examined any discrepancies between their individual measurements until a consensus was reached. The correlation between analyses conducted by both experimenters ranged from 0.92 to 1.00 across measures. For the three key words of each sentence, the acoustic measurements obtained were the fundamental frequency average (Hz), the fundamental frequency maximum (Hz), the duration (seconds), the intensity level average (dB), and the intensity level maximum (dB).

Table 1 provides the mean results for all key words in all sentences, by talker, for all five measurements. The mean fundamental frequency (F0) for all key words in all sentences differed by talker gender. Both female talkers had higher mean F0s (F1: 162.0 Hz and F2: 151.4 Hz), whereas both male talkers had lower mean F0s (M1: 110.6 Hz and M2: 90.6 Hz). The mean fundamental frequency maximum for all key words in all sentences also differed by talker gender in a similar manner. Both female talkers had higher maximum mean F0s (F1: 180.8 Hz and F2: 167.1 Hz), whereas both male talkers had lower maximum mean F0s (M1: 121.8 Hz and M2: 98.5 Hz). Conversely, for the mean duration for all key words in all sentences, there was little variation among all talkers (F1: 0.3 s, F2: 0.4 s, M1: 0.4 s, and M2: 0.3 s). The mean intensity for all key words in all sentences followed this pattern, as well with little variation among all talkers (F1: 57.5 dB, F2: 58.4 dB, M1: 58.3 dB, and M2: 57.0 dB). The mean intensity maximum for all key words in all sentences had little variation between all talkers (F1: 67.6 dB, F2: 71.0 dB, M1: 68.5 dB, and M2: 67.9 dB).

Figures 1-6 display differences in fundamental frequency, duration, and intensity between the unstressed and stressed sentence stimuli for each syntactic place condition through combined measurements for all talkers. These figures all contain box plots, where the dark line at the center of each boxplot marks the median value for the measurement at that syntactic place condition. Figures 1, 3, and 5 exhibit fundamental frequency, duration, and intensity for the unstressed sentence stimuli only. The X-ordinate signifies syntactic place, while the Y-ordinate displays one of the three measurements (in Hz, seconds, or dB, respectively). Figures 2, 4, and 6 exhibit fundamental frequency, duration, and intensity for each place condition for the stressed sentence stimuli. Each figure contains three panels, one for each place condition (e.g. ‘person’, ‘verb’, ‘object’). The X-ordinate of each panel signifies syntactic place condition, while the Y-ordinate displays one of the three measurements for all panels (in Hz, seconds, or dB).

Figure 1 shows the measured fundamental frequency for each of the three key words in the unstressed condition for all talkers combined. The ‘person’ words have the highest median fundamental frequency, while the ‘object’ words have the lowest. This follows Cooper et al. (1985) where fundamental frequency decreases throughout the production of single utterances. Variance in fundamental frequency is similar across all syntactic places.

Figure 2 displays the measured fundamental frequency for each of the three key words in the stressed conditions for all talkers combined. In the ‘person’ stressed condition (upper panel), the ‘person’ words have the highest median fundamental frequency, followed by the ‘verb’ words and then the ‘object’ words. Variance in fundamental frequency is greatest for the ‘person’ and ‘object’ words, whereas variance is minimal for the ‘verb’ words. In the ‘verb’ stress condition (middle panel), the ‘verb’ words have the highest median fundamental frequency, followed by the ‘person’ words and then the ‘object’ words. Variance in fundamental frequency is greatest for the ‘person’ and ‘verb’ words, whereas variance is more restricted for the ‘object’ words. In the ‘object’ stressed condition (lower panel), the ‘object’ words have the highest median fundamental frequency, followed by the ‘person’ and ‘verb’ words, which are essentially equal. Variance in fundamental frequency is greatest for the ‘object’ words, whereas variance is moderately reduced for the ‘person’ and ‘verb’ words.

A comparison of Figures 1 and 2 show that fundamental frequency increased for each syntactic place condition relative to other words in the sentence in the contrastive stress conditions. Cooper et al. (1985) had similar results, except for the first syntactic place condition.

They found little difference in fundamental frequency for the first word position between emphasized and non-emphasized conditions. We see this in the 'verb' and 'object' stressed conditions for the first word position, but not in the 'subject' stressed condition. Additionally, Cooper et al. (1985) found the greatest drop in fundamental frequency to occur between the first and second key words, regardless of focus location. We found this to only occur in the 'person' stressed condition. This difference is likely the result of sentence stimuli length and syntactic frame, where Cooper et al. (1985) had complex declarative sentences and we had simple declarative sentences. Our stimuli mimic the expectations of Fry (1958), where fundamental frequency is highest for the focus words in their corresponding stressed place condition.

Figure 3 illustrates the measured duration for each of the three key words in the unstressed condition with all talkers combined. The 'object' words have the highest median duration, followed by the 'person' and 'verb' words, which are effectively equivalent. Variance in duration is greatest for the 'object' words, whereas variance is reduced for the 'person' and 'verb' words.

Figure 4 shows the measured duration for each of the three key words in the stressed conditions with all talkers combined. In the 'person' stressed condition (upper panel), the 'object' words have the highest median duration, followed by the 'verb' and 'person' words, which are equivalent. Variance in duration is greatest for the 'object' words, whereas variance is much more restricted for the 'person' and 'verb' words. In the 'verb' stressed condition (middle panel), the 'verb' and 'object' words have the highest median duration, followed by the 'person' words. Similarly, variance in duration is greatest for the 'verb' and 'object' words, whereas variance is restricted for the 'person' words. Finally, in the 'object' stressed condition (bottom panel), the 'object' words by far have the highest median duration, followed by the 'person' and 'verb' words, which were similar. Variance in duration is greatest for the ‘object’ words, whereas variance is limited for the ‘person’ and ‘verb’ words.

A comparison of Figures 3 and 4 show that duration does not differ greatly between the unstressed and stressed conditions. This contrasts with Cooper et al. (1985) and Fry (1958), where the durations of key words are significantly greater when they are the focus of a sentence. We only see this for the ‘object’ stressed condition. Further, Cooper et al. (1985) found focus words at the end of sentences to have shorter durations than sentence-initial and sentence-medial focus words. We see the opposite for our sentence stimuli, where sentence-final words have the longest durations. As previously stated, these differences may be in part due to sentence stimuli length and syntactic frame.

Figure 5 displays the measured intensity for each of the three key words in the unstressed condition with all talkers combined. The ‘person’ words have the highest median intensity, while the ‘object’ words have the lowest. Intensity decreases from sentence-initial to sentence-final words. Variance in intensity is greatest for the ‘verb’ and ‘object’ words, whereas variance is more restricted for the ‘person’ words.

Figure 6 illustrates the measured intensity for each of the three key words in the stressed conditions with all talkers combined. In the ‘person’ stressed condition (upper panel), the ‘person’ words have the highest median intensity, followed by the ‘verb’ and ‘object’ words. Variance in intensity is greatest for the ‘verb’ and ‘object’ words, whereas variance is reduced for the ‘person’ words. In the ‘verb’ stressed condition (middle panel), the ‘verb’ words have the highest median intensity, followed closely by the ‘person’ words and then the ‘object’ words. Variance in intensity is greatest for the ‘verb’ and ‘object’ words, followed closely by the ‘person’ words. In the ‘object’ stressed condition (lower panel), the ‘person’ words and ‘verb’ words have an equivalent median intensity, followed by the 'object' words with the lowest median intensity. Variance in intensity was greatest for the 'person' and 'object' words, followed closely by the 'verb' words.

A comparison of Figures 5 and 6 show that intensity increased for each syntactic place condition relative to other words in the sentence in the stressed conditions. We found that intensity primarily increased for the focus word in their corresponding stressed syntactic place condition. Non-focus words were at similar levels in the stressed condition to the unstressed condition. This follows the expectations of Fry (1958).

b. Processed Stimuli

Sentence stimuli were processed using MATLAB (Mathworks, Natick MA) and converted via a 24-bit digital-to-analog converter. The stimuli were sent to a series of 16 band-pass filters. The temporal envelope of each band was extracted using a Hilbert transform. The resulting envelope was low-pass filtered using a fourth-order Butterworth filter. All 16 temporal envelopes were used to modulate a pure tone at the center frequency of each band, respectively (Oxenham & Kreft, 2014). A sine-wave carrier was chosen for two reasons: (1) sine-wave vocoded speech conveyed clearer periodicity cues; and (2) sine-wave vocoded speech was qualitatively more similar to the speech signal received by listeners with cochlear implants (Loebach, 2007).

There were three vocoded conditions, each using a separate low-pass cutoff frequency at the output of the 16 band-pass filters. These three cutoff frequencies were 50 Hz, 160 Hz, and 250 Hz. Oxenham and Kreft (2014) used a 50-Hz cutoff for their vocoded stimuli to eliminate periodicity and sideband cues. We used the same filter cutoff for the first condition to reduce the transmission of periodicity cues to listeners. In addition, we used two more higher-frequency cutoff filter conditions that would pass successively more periodic information to the listeners. These additional cutoff frequencies were selected based on the acoustic analysis of the talkers' productions of all stimuli. The 160 Hz cutoff was equivalent to the value of the two male talkers' productions for the highest fundamental frequencies in any conditions. Similarly, the 250 Hz cutoff was equivalent to the value of the two female talkers' productions for the highest fundamental frequencies in any conditions.

c. Participants

The participants were 33 adults with normal hearing (22 female, age range 18-35, mean = 23.3 years, standard deviation = 3.7 years). Each participant passed a hearing screening bilaterally at 20 dB HL for the frequencies of 0.5, 1, 2, and 4 kHz. Participants were each assigned to one of four groups. The first group received the natural, unprocessed stimuli. The second, third, and fourth groups received processed (vocoded) stimuli, with low-pass cutoff filters of 50 Hz, 160 Hz, or 250 Hz, respectively.

Table 2 contains the participant characteristics for the unprocessed and processed groups. The first group had 9 participants (7 female, age range 18-35, mean = 24.5 years, standard deviation = 5.9 years). The second, third, and fourth groups had 24 participants in total (15 female, age range 19-29, mean = 22.8 years, standard deviation = 2.4 years). The groups presented with processed stimuli were balanced, with eight participants in each. Appendices B and C provide specific demographic information for each participant, by unprocessed and processed group conditions, respectively. According to IRB standards, each participant was asked to read and sign a consent form and a HIPAA form before participating.

d. Procedure

Participants sat in a double-walled sound treated room. Stimuli were generated using the 24-bit output from a Lynx 122 sound card and a Benchmark DAC1 digital-to-analog converter. Stimuli were presented at an average of 60 dB SPL through two Bose 8-ohm bookshelf speakers connected to a Tucker-Davis HB6 headphone buffer module. The speakers were placed at 45 degrees azimuth to the participants. A computer, keyboard, and mouse were used to complete the tasks. Verbal instruction was provided for each condition and the experimenter initiated each condition for all participants. There were three conditions: (1) sentence recognition, (2) training for contrastive stress, and (3) contrastive stress. All responses were collected using MATLAB (Mathworks, Natick MA) scripts through graphic use interfaces (GUIs).

i. Condition 1: Sentence Recognition

The first condition was a sentence recognition task to measure sentence intelligibility. Via experimenter instruction, participants were asked to listen to the sentence stimuli and type each sentence after its presentation. The experimenter initiated the task and participants controlled the pace by clicking play to hear each sentence. Participants were instructed to save their response when they were satisfied with their answer.

Each participant was randomly assigned one of four corpuses of sentences. Each corpus was composed of 72 non-stressed sentences (18 sentences per talker) randomly selected from the combined 576 sentence stimuli of the four talkers. Appendices D and E provide the corpus assignment for each participant, for both the unprocessed and processed group conditions, respectively. A sentence recognition score, in percent correct words, was obtained for each participant. Responses were autoscored for the three key words of each sentence ('person', ‘verb’, ‘object’). Responses were then hand scored in the presence of errors. Hand scoring was done to ensure that spelling and syntactic errors were not incorrectly identified as being wrong.

ii. Condition 2: Training for Contrastive Stress

The second condition was a training task for the judgment of contrastive stress. By experimenter instruction, participants were asked to listen to each stimulus and select the word they believed was stressed within each sentence from the choices on the computer monitor. If they believed there was no stress within a sentence, participants were asked to select the ‘no stress’ option. This was a forced-choice task, with four options: ‘person’, ‘verb’, ‘object’, or ‘no stress’. After selecting their response, participants were provided with feedback on the correct response.

Each participant heard 48 sentences. For this task, a new female talker (F3) produced the sentences in the same manner as the four experimental talkers. F3 was selected for her clarity and naturalness in speaking. Further, it was intended that this talker be different than those used for the final contrastive stress experiment. A training score, in percent correct, was obtained for each participant.

iii. Condition 3: Contrastive Stress

The third condition was a contrastive stress experimental test. Similar to the training task, via experimenter instruction, participants were asked to listen to each stimulus and select the word they believed was stressed within each sentence from the choices on the computer monitor. If they believed there was no stress within a sentence, participants were asked to select the ‘no stress’ option. This was a forced-choice task, with four options: ‘person’, ‘verb’, ‘object’, or ‘no stress’. Participants were not provided with feedback on the correct response.

Each participant was presented with 144 sentences spoken by each of the four talkers (F1, F2, M1, M2). Talkers were presented in a randomized order for each listener. A total of 576 sentences were presented to each participant. Appendices D and E provide the talker order assignment for each participant, for both the unprocessed and processed group conditions, respectively. A contrastive stress test score, in percent correct, was obtained for each participant for each talker.

III. Results

a. Sentence Recognition

i. Unprocessed Speech Condition

Table 3 shows the mean sentence recognition score for both the unprocessed and processed group conditions. The unprocessed group had a near-perfect mean recognition score of 99.7% (standard deviation = 0.5%). Figure 7 displays the median and interquartile range of the sentence recognition scores (in percent correct) for participants in the unprocessed condition. There is little within group variation of sentence recognition scores. Appendix F provides individual participant sentence recognition scores.

ii. Processed Speech Conditions

All three groups of participants who received the processed speech conditions showed near-perfect sentence recognition scores. For the processed condition with a low-pass filter cutoff of 50 Hz, the mean score was 97.2% (standard deviation = 4.4%). For the low-pass filter cutoff condition of 160 Hz, the mean score was 97.9% (standard deviation = 2.0%). For the low-pass filter cutoff condition of 250 Hz, the mean score was 97.9% (standard deviation = 1.3%). Appendix G provides individual participant sentence recognition scores.

Figure 8 displays the median and interquartile ranges of percent correct sentence recognition as a function of filter cutoff. The medians for all low-pass filter cutoff conditions are relatively equal. Similarly, the variance within each group is comparable among all groups, as demonstrated by the placement of the box limits for the 25^{th} and 75^{th} percentiles for each low-pass filter cutoff condition. Within group variation was minimal in all low-pass filter cutoff conditions. Both the 50 Hz and 160 Hz low-pass filter cutoff conditions had one outlier each at the lower 10^{th} percentile.

A one-way analysis of variance (ANOVA) was completed to examine the between subject factor of filter cutoff frequency for sentence recognition. No significant difference was found among the three filter groups in sentence recognition score averages ($F(2,21) = 1.6$, p-value = 0.2). A test of multiple comparisons using Tukey HSD showed that each filter condition was not significantly different from the other two filter conditions. Thus, the speech recognition scores for the 50 Hz filter cutoff condition were not significantly different from the scores for the 160 Hz and 250 Hz filter cutoff conditions (p-values = 0.9 and 0.9, respectively). Similarly, the speech recognition scores for the 160 Hz filter cutoff condition were not significantly different from the scores for the 50 Hz and 250 Hz filter cutoff conditions (p-values = 0.9 and 1.0, respectively). Lastly, the speech recognition scores for the 250 Hz filter cutoff condition were not significantly different from the scores for the 50 Hz and 160 Hz filter cutoff conditions (p-values = 0.9 and 1.0, respectively). Stimuli were highly intelligible for all filter conditions.

c. Training for Contrastive Stress

i. Unprocessed Speech Condition

Table 4 shows the mean contrastive stress training score for both the unprocessed and processed group conditions. The unprocessed group had a contrastive stress training score of 91.2% (standard deviation = 7.9%). Appendix H provides individual participant contrastive stress training scores.

ii. Processed Speech Condition

All three groups of participants who received the processed speech conditions show varying levels of contrastive stress training scores. For the processed condition with a low-pass filter cutoff of 50 Hz, the mean score was 60.4% (standard deviation = 11.0%). For the low-pass filter cutoff condition of 160 Hz, the mean score was 57.6% (standard deviation = 12.9%). For the low-pass filter cutoff condition of 250 Hz, the mean score was 64.3% (standard deviation = 9.7%). Appendix I provides individual participant contrastive stress training scores.

Figure 9 displays the median and interquartile ranges of percent correct contrastive stress training scores as a function of filter cutoff. The 160 Hz low-pass filter cutoff condition has the highest median score, followed closely by the 250 Hz low-pass filter cutoff condition. The amount of within group variation is similar across all low-pass filter cutoff conditions. Although, as demonstrated by the placement of the box limits for the 25th and 75th percentiles for each low-pass filter cutoff condition, the 50 Hz low-pass filter cutoff condition has greater variance in the upper fence, the 160 Hz low-pass filter cutoff condition has greater variance in the lower fence, and the 250 Hz low-pass filter cutoff condition has equivalent variance in both fences.

d. Contrastive Stress Experimental Test

i. Unprocessed Speech Condition

Table 5 shows the mean contrastive stress experimental test score for both the unprocessed and processed group conditions. The unprocessed group had a mean contrastive stress experimental test score of 93.2% (standard deviation = 5.2%). Appendix J provides individual participant contrastive stress experimental test scores. Appendix K provides mean contrastive stress experimental test scores for each talker individually by all syntactic place conditions.

ii. Processed Speech Condition

All three groups of participants who received the processed speech conditions had wide ranges of contrastive stress experimental test scores. For the low-pass filter cutoff condition of 50 Hz, the mean score was 73.3% (standard deviation = 7.1%). For the low-pass filter cutoff condition of 160 Hz, the mean score was 68.6% (standard deviation = 8.1%). Finally, for the low-pass filter cutoff condition of 250 Hz, the mean score was 82.3% (standard deviation = 6.3%). Appendix L provides individual participant contrastive stress experimental test scores. Appendices M, N, and O provide mean contrastive stress experimental test scores for each talker individually by all syntactic place conditions for each low-pass filter cutoff condition, respectively.

Figure 10 displays the median and interquartile ranges of percent correct contrastive stress experimental test scores as a function of filter cutoff. The 250 Hz low-pass filter cutoff condition has the highest median score, followed by the 50 Hz low-pass filter cutoff condition and then the 160 Hz low-pass filter cutoff condition. The amount of within group variation is similar across all low-pass filter cutoff conditions. As demonstrated by the placement of the box limits for the 25th and 75th percentiles for each low-pass filter cutoff condition, the 160 Hz low-pass filter cutoff condition has greater variance in the upper fence, whereas the 50 Hz and 250 Hz low-pass filter cutoff conditions have relatively equivalent variance in both fences.

Figure 11 shows a scatterplot of participant contrastive stress experimental test scores as a function of sentence recognition scores for the low-pass filter cutoff conditions. Except for two outliers, sentence recognition scores for all participants are perfect or near perfect. In contrast, we see a wide distribution for contrastive stress experimental test scores. Participants in the 250 Hz low-pass filter cutoff condition group appeared to have higher scores on average than participants in the 50 Hz and 160 Hz low-pass filter cutoff condition groups.

Figures 12-15 show the median and interquartile ranges of percent correct contrastive stress experimental test scores for talkers F1, F2, M1, and M2, respectively. Each figure contains three panels: the first for the 50 Hz low-pass filter cutoff condition, the second for the 160 Hz low-pass filter cutoff condition, and the third for the 250 Hz low-pass filter cutoff condition. Each panel contains four box plots, where the dark line at the center of each boxplot marks the median value for the measurement at that syntactic place condition ('none', 'person', 'verb', 'object'). The X-ordinate signifies syntactic place, while the Y-ordinate displays contrastive stress score (in percent correct).

Figure 12 displays the median and interquartile ranges of percent correct contrastive stress experimental test scores for talker F1 by filter cutoff as a function of sentence stress condition ('none', 'person', 'verb', 'object'). For the 50 Hz low-pass filter cutoff condition, 'object' has the highest median value, followed closely by 'person', then 'none' and 'verb'. Variance in percent correct contrastive stress score is greatest for 'none' and 'verb', with greater variance in the lower fences for both. Variance is minimal for 'person' and 'object', with equivalent variance in both fences for 'person' and greater variance in the upper fence for 'object'. For the 160 Hz low-pass filter cutoff condition, 'object' and 'person' have the highest median values, followed by 'verb' and then 'none'. Variance is greatest for 'none' and 'verb', with equal variance in both fences for 'none' and 'verb'. Variance is more restricted for 'person' and 'object', with greater variance in the lower fences for both 'person' and 'object'. For the 250

Hz low-pass filter cutoff condition, ‘person’ has the highest median value, followed closely by ‘object’, then ‘none’ and ‘verb’. Variance is greatest for ‘verb’, with greater variance in the lower fence. Variance is also great for ‘none’, with approximately equivalent variance in both fences. Variance is more limited for ‘person’ and ‘object’, with greater variance in the lower fences for both. For both the 50 Hz and 250 Hz low-pass filter cutoff conditions, an outlier is present at the lower 10th percentile for ‘person’.

Figure 13 displays the median and interquartile ranges of percent correct contrastive stress experimental test scores for talker F2 by filter cutoff as a function of sentence stress condition (‘none’, ‘person’, ‘verb’, ‘object’). For the 50 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed closely by ‘none’, ‘person’, and then ‘object’. Variance in percent correct contrastive stress score is roughly equivalent for ‘none’, ‘person’, and ‘verb’, with greater variance in the lower fences for both ‘none’ and ‘person’ and equal variance in both fences for ‘verb’. Variance is more restricted for ‘object, with greater variance in the lower fence. For the 160 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed by ‘person’, ‘none’, and ‘verb’. Variance is greatest for ‘none’ and ‘verb’, with greater variance in the lower fence for ‘none’ and equivalent variance in both fences for ‘verb’. Variance is smaller for ‘person’ and ‘object’, with greater variance in the lower fences for ‘person’ and ‘object’. For the 250 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed closely by ‘person’ and ‘none’, then ‘verb’. Variance is greatest for ‘none’, ‘verb’, and ‘object’, with greater variance in the lower fences for all three. Variance is more restricted for ‘person’, with greater variance in the lower fence. For the 250 Hz low-pass filter cutoff condition, an outlier is present at the lower 10th percentile for ‘person’ and ‘verb’.

Figure 14 displays the median and interquartile ranges of percent correct contrastive stress experimental test scores for talker M1 by filter cutoff as a function of sentence stress condition (‘none’, ‘person’, ‘verb’, ‘object’). For the 50 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed by ‘verb’, then ‘person’ and ‘none’. Variance in percent correct contrastive stress score is greatest for ‘none’, with greater variance in the lower fence. Variance is also high for ‘person’ and ‘verb’, with greater variance in the upper fence for ‘person’ and equal variance in both fences for ‘verb’. Variance is more restricted for ‘object’, with greater variance in the upper fence. For the 160 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed closely by ‘person’, then ‘verb’ and ‘none’. Variance is greatest for ‘none’ and ‘verb’, with greater variance in the upper fence for ‘none’ and greater variance in the lower fence for ‘verb’. Variance is limited for ‘person’ and ‘object’, with greater variance in the lower fence for both. For the 250 Hz low-pass filter cutoff condition, ‘person’ has the greatest median value, followed closely by ‘person’, then ‘verb’ and ‘none. Variance is greatest for ‘none’ and ‘verb’, with greater variance in the upper fence for ‘none’ and equal variance in both fences for ‘verb’. Variance is more restricted for ‘person’ and ‘object’, with greater variance in the lower fence for ‘person’ and equal variance in both fences for ‘object’. For both the 160 Hz and 250 Hz low-pass filter cutoff conditions, an outlier is present at the lower 10^{th} percentile for ‘person’. For the 250 Hz low-pass filter cutoff condition, an outlier is present at the lower 10^{th} percentile for ‘object’.

Figure 15 displays the median and interquartile ranges of percent correct contrastive stress experimental test scores for talker M2 by filter cutoff as a function of sentence stress condition (‘none’, ‘person’, ‘verb’, ‘object’). For the 50 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed closely by ‘verb’, ‘person’, and ‘none’. Variance in percent correct contrastive stress score is greatest for ‘none’, with greater variance in the upper fence. Variance is more restricted for ‘person’, ‘verb’, and ‘object’, with greater variance in the lower fence for all. For the 160 Hz low-pass filter cutoff condition, ‘object’ has the greatest median value, followed closely by ‘person’, then ‘verb’ and ‘none’. Variance is greatest for ‘none’, ‘person’, and ‘verb’, with equal variance in both fences for ‘none’ and greater variance in the lower fence for both ‘person’ and ‘verb’. Variance is limited ‘object’, with greater variance in the upper fence. For the 250 Hz low-pass filter cutoff condition, ‘verb’ and ‘object’ have equivalently great median values, followed by ‘none’ and ‘person’. Variance is greatest for ‘person’, with greater variance in the upper fence for ‘person’. Variance is more limited in ‘none’ and ‘verb’, with greater variance in the lower fence for both. Variance is even more restricted for ‘object’, with greater variance in the lower fence for ‘object’. For the 250 Hz low-pass filter cutoff condition, an outlier is present at the lower 10^{th} percentile for ‘none’.

A three-way analysis of variance (ANOVA) with repeated measures was completed with one between factor and two within factors for correct contrastive stress response. The between-participant factor consisted of the three group conditions: stimuli with a low-pass cutoff at 50 Hz, stimuli with a low-pass cutoff at 160 Hz, and stimuli with a low-pass cutoff at 250 Hz. The two within-participant factors were the talker and the place of stress, each with four levels of repeated measures. The talkers were: F1, F2, M1, and M2. The places of stress were: ‘person’, ‘verb’, ‘object’, and ‘no stress’. All participants heard stimuli from all four talkers and in all four stress conditions. Three significant main effects were found for filter, talker, and place of stress. A significant effect was also found for the interaction of talker by place.

The first significant main effect was for filter cutoff frequency ($F(2,21) = 5.0$, p-value = 0.02). A test of multiple comparisons using Tukey HSD shows that the contrastive stress experimental test scores for the 250 Hz low-pass filter condition are significantly different from the 50 Hz and 160 Hz low-pass filter conditions (p-values = 0.04 and 0.02, respectively). This test also shows that the contrastive stress experimental test scores for the 50 Hz filter cutoff condition and the 160 Hz filter cutoff condition are not significantly different from each other (p-value = 1.0). Refer to Table 5 for the filter condition group means and standard deviations.

A test of within-subject effects using the Greenhouse-Geisser correction shows a significant main effect for talker ($F(2.4,51.1) = 37.6$, p-value < 0.001). For talkers F1, F2, and M1, we see mean contrastive stress experimental test scores of 74.6% (standard error = 1.7%), 75.6% (standard error = 1.9%), and 73.5% (standard error = 1.8%), respectively. For talker M2, we see a mean contrastive stress experimental test score of 89.0% (standard error = 1.1%). These mean scores are derived from the individual participant scores across all low-pass filter condition groups.

A test of within-subject effects using the Greenhouse-Geisser correction shows a significant main effect for syntactic place ($F(2.1, 43.6) = 21.2$, p-value < 0.001). For the no stress condition, we see a mean contrastive stress experimental test score of 67.9% (standard error = 2.8%). For person, we see a mean score of 81.8% (standard error = 2.1%). For verb, we see a mean score of 72.2% (standard error = 2.9%). Lastly, for object, we see a mean score of 90.9% (standard error = 1.1%). These mean scores are derived from the individual participant scores across all low-pass filter condition groups.

A test of within-subject effects using the Greenhouse-Geisser correction shows a significant main effect for the interaction of talker by place ($F(5.3,111.5) = 16.0$, p-value $<$ 0.0001). For talker F1, the mean contrastive stress test scores for ‘none’, ‘person’, ‘verb’, and ‘object are 63.8% (standard deviation = 18.5%), 87.0% (standard deviation = 12.9%), 58.7% (standard deviation = 19.7%), and 89.1% (standard deviation = 9.0%), respectively. For talker F2, the mean contrastive stress test scores for ‘none’, ‘person’, ‘verb’, and ‘object are 71.3% (standard deviation = 19.9%), 77.1% (standard deviation = 14.9%), 67.5% (standard deviation = 19.7%), and 86.3% (standard deviation = 10.2%), respectively. For talker M1, the mean contrastive stress test scores for ‘none’, ‘person’, ‘verb’, and ‘object are 55.0% (standard deviation = 20.7%), 77.8% (standard deviation = 20.5%), 69.3% (standard deviation = 16.7%), and 92.0% (standard deviation = 5.1%), respectively. For talker M2, the mean contrastive stress test scores for ‘none’, ‘person’, ‘verb’, and ‘object are 81.3% (standard deviation = 14.8%), 85.2% (standard deviation = 9.1%), 93.2% (standard deviation = 8.6%), and 96.4% (standard deviation = 3.9%), respectively. No significant effects were found for the interactions of talker by filter ($F(4.9,51.1) = 1.0$, p-value = 0.4), place by filter ($F(4.2,43.6) = 2.2$, p-value = 0.1), or talker by place by filter ($F(10.6,111.5) = 1.1$, p-value = 0.4).

IV. Discussion and Conclusion

The results of the sentence recognition task for both the unprocessed and processed speech group conditions indicate that our stimuli were highly intelligible and that the task was clear to participants. Across all groups, participants were able to perform the sentence recognition task near ceiling levels. Similarly, high performance by participants in the unprocessed speech group condition for the contrastive stress training task indicates that the contrastive stress training task was clear to participants and that contrastive stress is perceived in the natural unprocessed speech sentence stimuli.

Participants had lower performance in the three processed speech group conditions for the contrastive stress training task, indicating greater difficulty in the perception of contrastive stress in the presence of reduced periodicity cues. Performance by the three processed speech condition groups was substantially above chance, which in a forced-choice task with four options is 25%. This further indicates clarity of the contrastive stress training task.

Participants' high performance in the unprocessed speech group condition for the contrastive stress experimental task supports the clarity of the contrastive stress task and the perception of contrastive stress in natural unprocessed stimuli. We opted to not include participants in the unprocessed group in the analysis of variance. Instead we analyzed the differences among the processed speech group conditions because of the qualitative differences between unprocessed and processed speech. Rather, the results of the participants in the unprocessed speech group condition are being used to signify the legitimacy of the contrastive stress experimental test in the actual perception of contrastive stress.

We originally hypothesized that there would be a steady improvement in performance on the contrastive stress experimental task as a function of filter cutoff. Although we found a significant main effect of filter frequency cutoff, it was not as expected. The 50 Hz and 160 Hz low-pass filter cutoff conditions resulted in levels of performance that were not significantly different, whereas the results for the 250 Hz low-pass filter cutoff condition were significantly better than those for both 50 Hz and 160 Hz. This suggests that a sizeable increase in the availability of periodicity cues is necessary to see improved performance in the perception of contrastive stress through vocoded speech.

We initially hypothesized that there would be minimal individual talker effects, where performance by participants would vary little across talker conditions. Given the results of the acoustic analysis, we did not expect any one talker to elicit substantially different results from the others. We found that participants' performance for talker M2 was unlike that for talkers F1, F2, and M1. Participants had a mean contrastive stress experimental test score in the high-80s for M2 for all participants combined across the three processed group conditions, whereas F1, F2, and M1 elicited mean performance scores in the mid-70s.

Our acoustic analysis of the sentence stimuli showed that talker M2 had fewer variations in fundamental frequency for each of the sentence stress conditions, especially when compared to the two female talkers. We believe that low variability may relate to better performance by participants for talker M2. Further, due to the predictable nature and high intelligibility of our stimuli, we suggest that the differences in performance by talker may be the result of subtle speaker variations that did not show up in the spectral and temporal analyses that we completed; these led to disparities in task difficulty by talker. It has become common practice to use multiple talkers for both experimental tasks and recorded clinical tests and to aggregate the results across talker. It is less common to analyze the task/test results by talker. Our results suggests future studies may need to pay more attention to the specific characteristics of multiple talkers in and their impact on the perception of suprasegmental cues.

We did not anticipate a significant effect for syntactic place of stress. We hypothesized that we would expect to see better performance for sentence-initial and sentence-medial words if an effect were to be found for place of stress. This is based on the findings from Cooper et al. (1985), where the intonation curve showed greater fundamental frequency cues for contrastive stress at the initial and medial portions of utterances. Rather, we found that performance was greatest for 'object' words, followed by 'person', 'verb', and 'none' words. Participants had a mean contrastive stress experimental test score in the low-90s combined across the three processed group conditions for the 'object' stressed condition, for all participants. The 'person', 'verb', and 'none' stressed conditions resulted in mean scores in the low-80s, low-70s, and high-60s, respectively. These mean scores are all well above chance, suggesting that contrastive stress was perceived in all place conditions. Further, in the no stress sentence condition, participants were able to correctly identify the absence of stress with above chance performance. We believe though, that the relatively low mean score for the no stress condition, compared to other sentence stress condition mean scores, was the result of participants looking for stress where there was none.

We also did not anticipate a significant interaction for talker by place. We had hypothesized that there would be minimal variation in participants contrastive stress experimental test performance across talkers, as well as minimal variation in performance across sentence stress conditions. We found performance by participants was better for the 'person' and 'object' sentence stress conditions for talkers F1, F2, and M1, with lower performance for the ‘none’ and ‘verb’ conditions. For talker M2, we found performance by participants was best for ‘object’, then ‘verb’, ‘person’, and ‘none’. Once again, we relate this to speaker differences, where speakers convey different cues in varying manners.

There were no significant effects for the interaction of talker by filter. We had originally hypothesized that there would be a talker effect for gender by filter cutoff. We expected to see high performance by participants in the contrastive stress experimental test across all talkers for the 250 Hz cutoff filter, assuming all fundamental frequency information would be provided for all talkers. For the 160 Hz cutoff filter, we only expected to see high performance by participants for the male talkers, assuming all fundamental frequency information would be provided for the male talkers and not for the female talkers. Lastly, for the 50 Hz cutoff filter, we did not expect to see high performance by participants for any talker, due to the reduced periodicity cues for all talkers. Rather, we saw steady performance by participants between the 50 Hz and 160 Hz filter cutoff conditions, with no marked increases in performance by participants between the two filter conditions across all takers. Performance for the 250 Hz cutoff filter saw marked increases in performance by participants across all talkers. There were no talker effects for gender by filter cutoff. Across all filter cutoff conditions, we see similar performance by participants across talkers F1, F2, and M1. The only talker to show any marked differences is talker M2, where performance was higher than all other talkers across all filter conditions.

V. Conclusion

We sought to examine the effect of providing successively more periodicity cues have on the perception of contrastive stress. We found that sizeable increases in the availability of periodicity cues are necessary to see improved performance in the perception of contrastive stress. We also found that talker variations, as well as place of stress, play a role in participant variation in performance for the contrastive stress experimental test.

These results may serve to explain why cochlear implant users have had such variable performance in other studies of suprasegmental perception. Based on their devices, the signal processing algorithm used, and the effects of programming. CI users are likely to receive very different periodicity cues that may enhance or reduce their ability to use fundamental frequency cues that signal stress, intonation, and gender. Temporally based cues such as duration and intensity may not vary the same way and may have less of an impact.

Further, due to the variation in performance by participants across all talkers, there is the need to continue research in how individual speakers influence the perception of contrastive stress and other suprasegmental cues, especially with restricted cues. We suggest the continued use of larger samples of multiple talkers for studying contrastive stress and other suprasegmental cues in future research in which their individual speaking characteristics are tested. Additionally, more research is needed in how fundamental frequency truly influences contrastive stress perception. Watson and Schlauch (2008) studied the effect of fundamental frequency on intelligibility with flattened intonation contours on speech stimuli. Because sentences with flattened intonation curves yield poorer intelligibility scores, the use of flattened intonation contours on speech stimuli produced with contrastive stress may provide additional insight into the role fundamental frequency plays in the perception of contrastive stress for NH listeners with vocoded speech and for listeners with CIs.

VI. Tables

Table 1 Acoustic analysis measurement averages for all key words in all sentences, by talker.

	Talker			
	F1	F2	M1	M2
Fundamental Frequency Average (Hz)	162.03	151.44	110.57	90.58
Fundamental Frequency Max (Hz)	180.83	167.11	121.76	98.51
Duration (seconds)	0.34	0.39	0.37	0.32
Intensity Average (dB)	57.48	58.38	58.27	57.03
Intensity Max (dB)	67.57	70.96	68.52	67.90

Table 2 Group demographics, unprocessed and processed conditions.

	Unprocessed	Processed
n	9	24
Female	7	15
Male	2	9
Mean Age (years)	24.51	22.81
Standard Deviation (years)	5.91	2.38

Table 3 Sentence recognition score averages for the unprocessed and processed speech group conditions, in percent correct words.

	Unprocessed	Processed (by low-pass filter cutoff)		
		50 Hz	160 Hz	250 Hz
Mean Score (percent)	99.69	97.22	97.92	97.86
Standard Deviation (percent)	0.47	4.41	1.96	1.33

Table 4 Training for contrastive stress score averages for the unprocessed and processed speech group conditions, in percent correct.

	Unprocessed	Processed (by low-pass filter cutoff)		
		50 Hz	160 Hz	250 Hz
Mean Score (percent)	91.20	60.42	57.55	64.32
Standard Deviation (percent)	7.92	10.97	12.94	9.74

Table 5 Contrastive stress experimental test score averages for the unprocessed and processed speech group conditions, in percent correct.

	Unprocessed	Processed (by low-pass filter cutoff)		
		50 Hz	160 Hz	250 Hz
Mean Score (percent)	93.23	73.29	68.60	82.34
Standard Deviation (percent)	5.19	7.10	8.11	6.31

VII. Figures

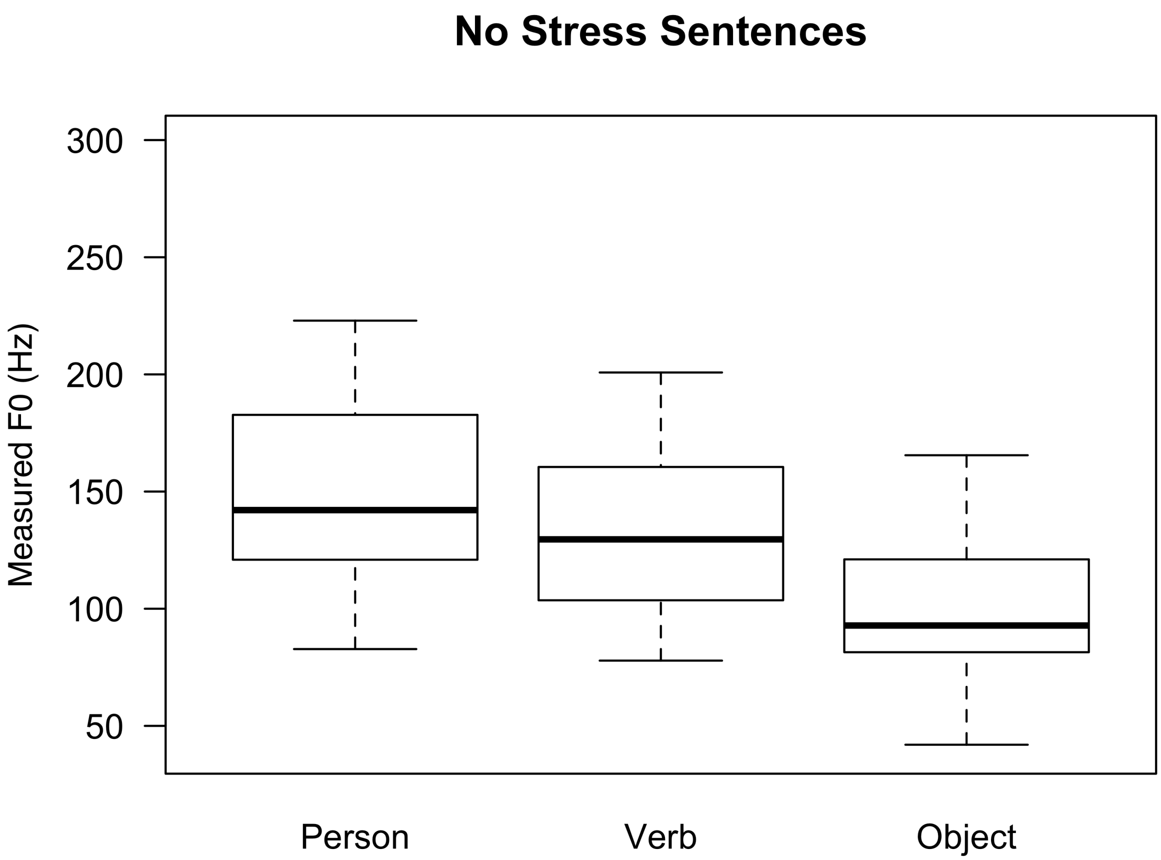

Figure 1 Box plot of measured fundamental frequency (F0) for each of the three key words in the no stress sentence condition, all talkers combined. The central lines represent the median values and the box limits are the 25^{th} and 75^{th} percentiles. The lower fence is from the 10^{th} to 25^{th} percentiles and the upper fence is from the 75^{th} to 90^{th} percentiles.

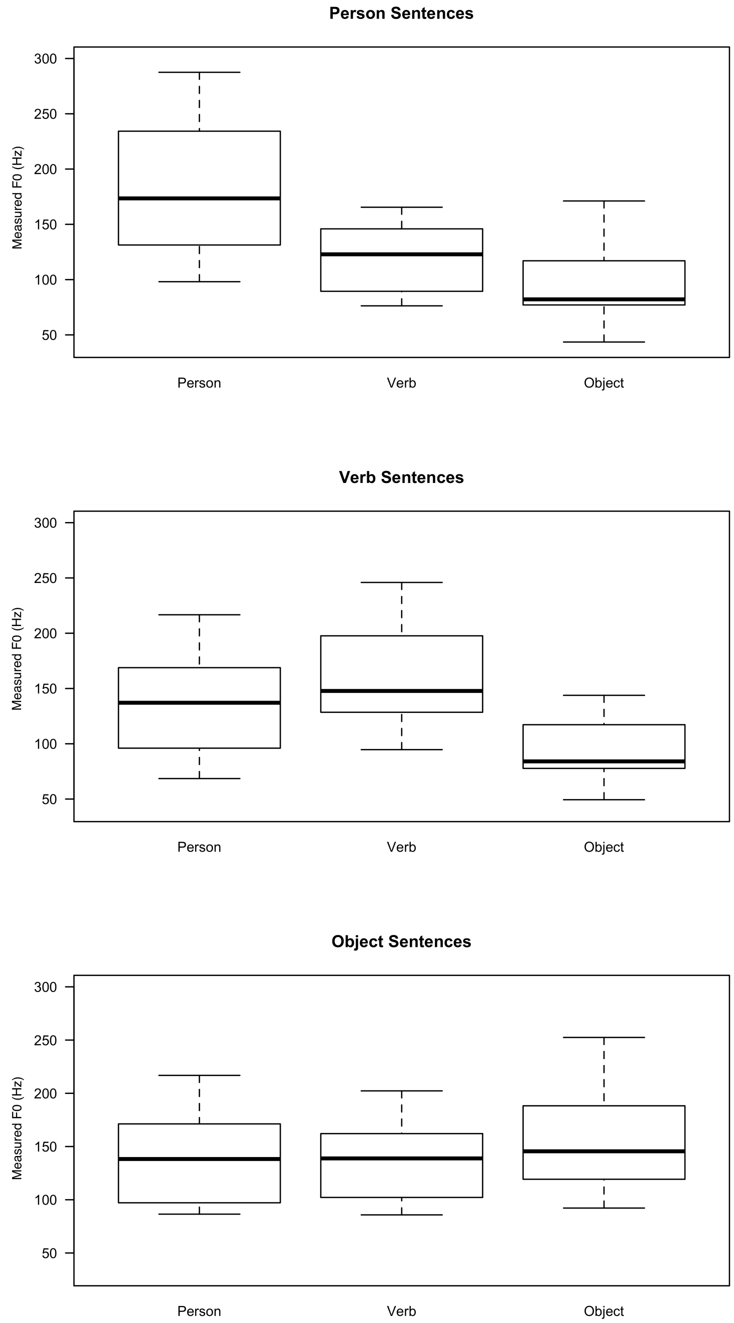

Figure 2 Box plot of measured fundamental frequency (F0) for each of the three key words in the stressed sentence conditions, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles.

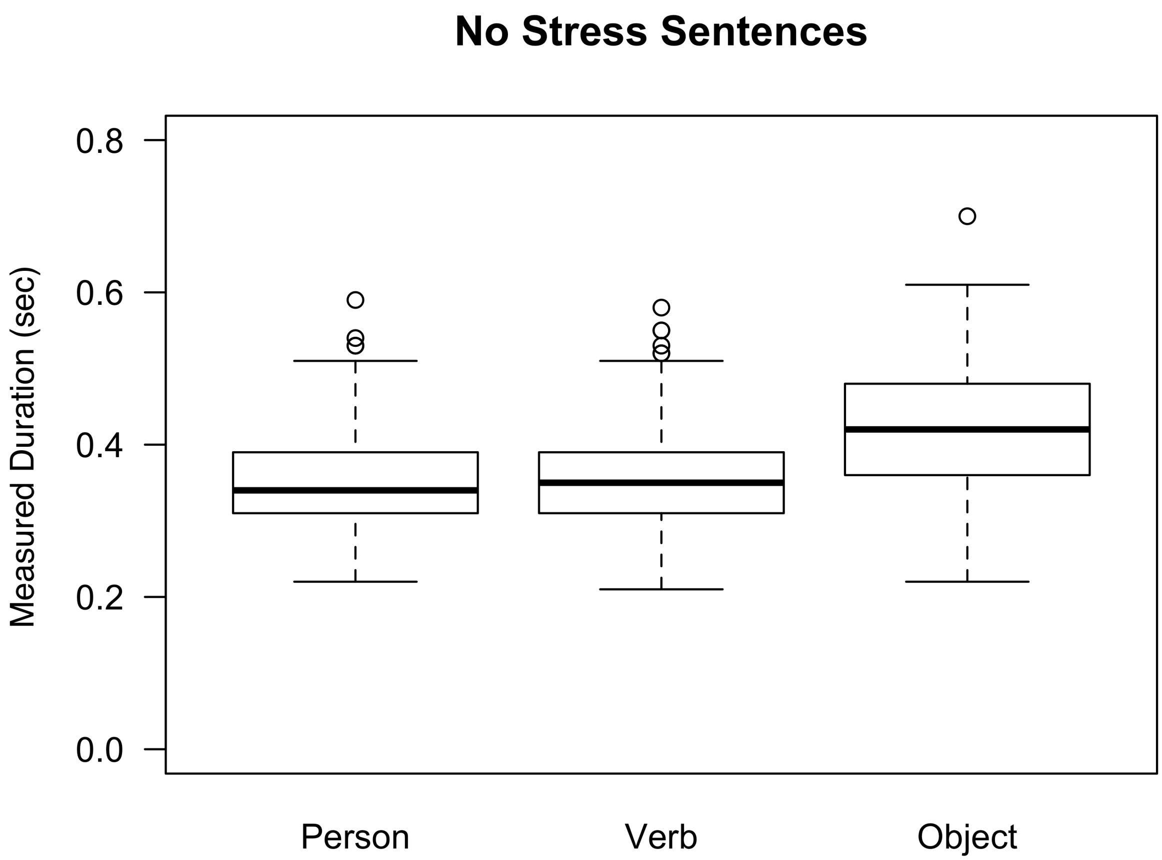

Figure 3 Box plot of measured duration for each of the three key words in the no stress sentence condition, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the upper 10th percentiles.

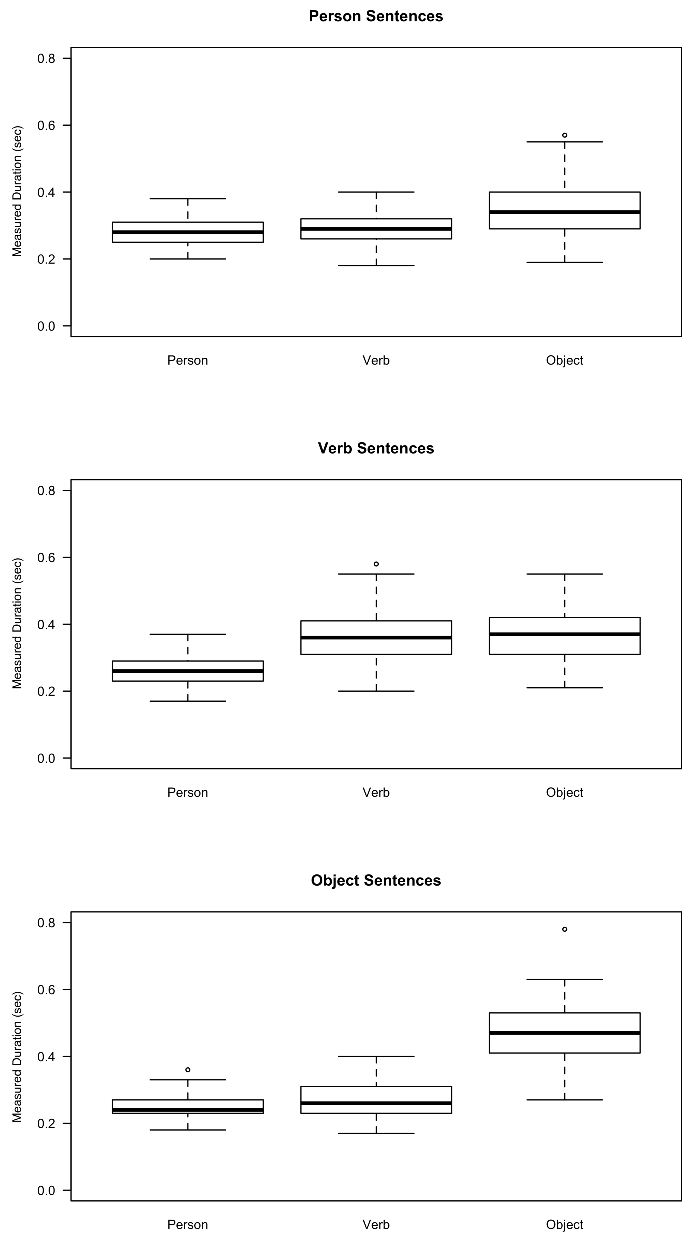

Figure 4 Box plot of measured duration for each of the three key words in the stressed sentence conditions, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the upper 10th percentiles.

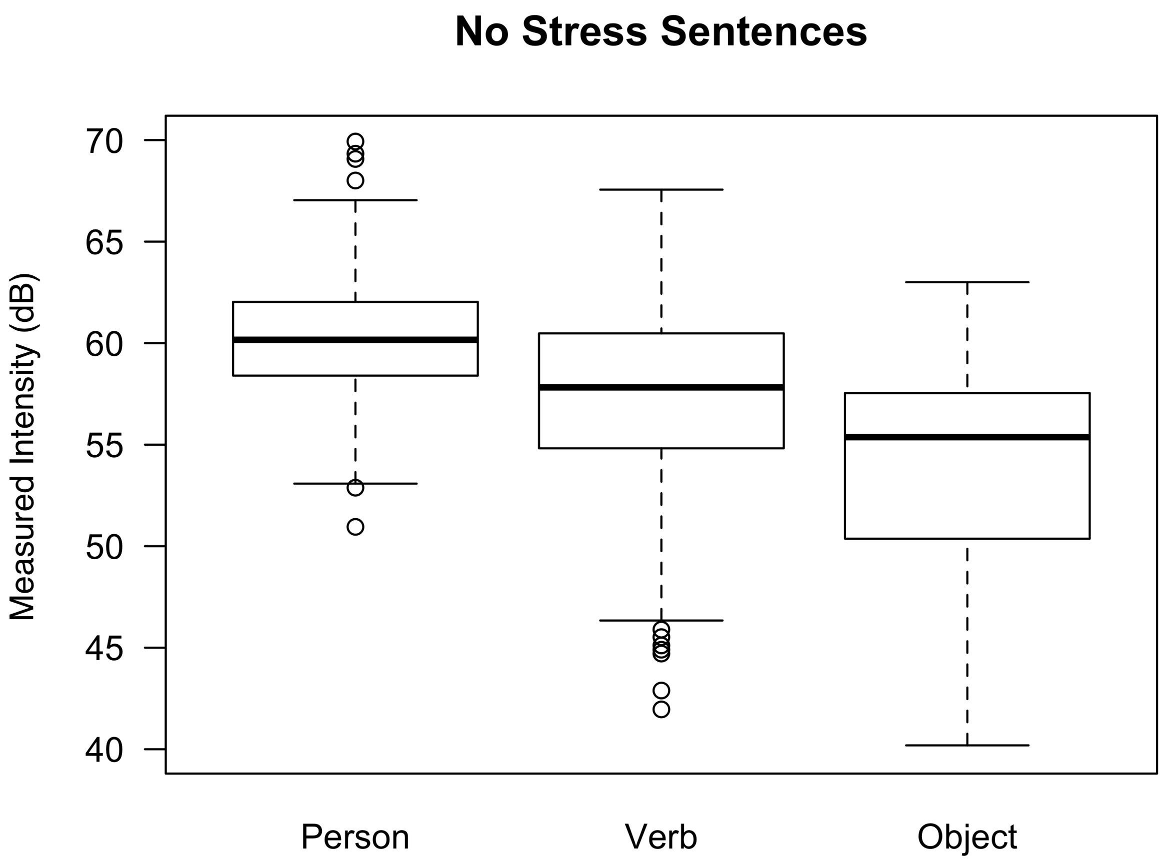

Figure 5 Box plot of measured intensity for each of the three key words in the no stress sentence condition, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the upper and lower 10th percentiles.

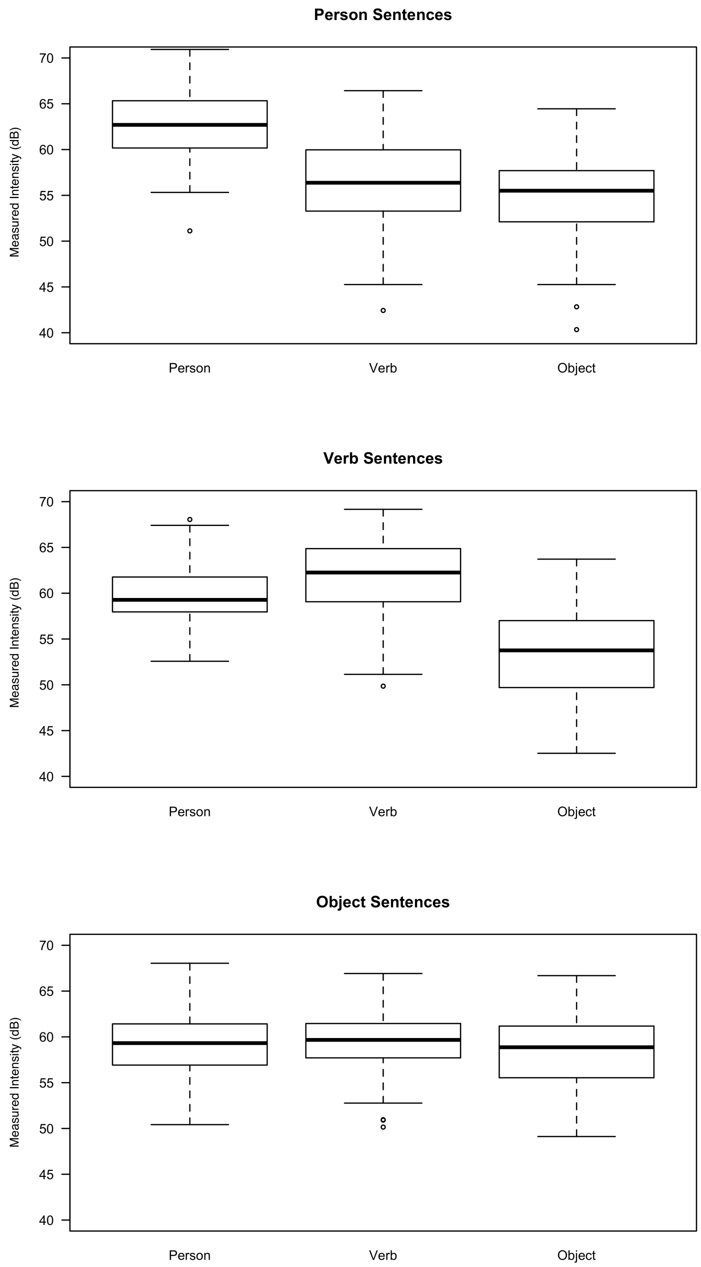

Figure 6 Box plot of measured intensity for each of the three key words in the stressed sentence conditions, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the upper and lower 10th percentiles.

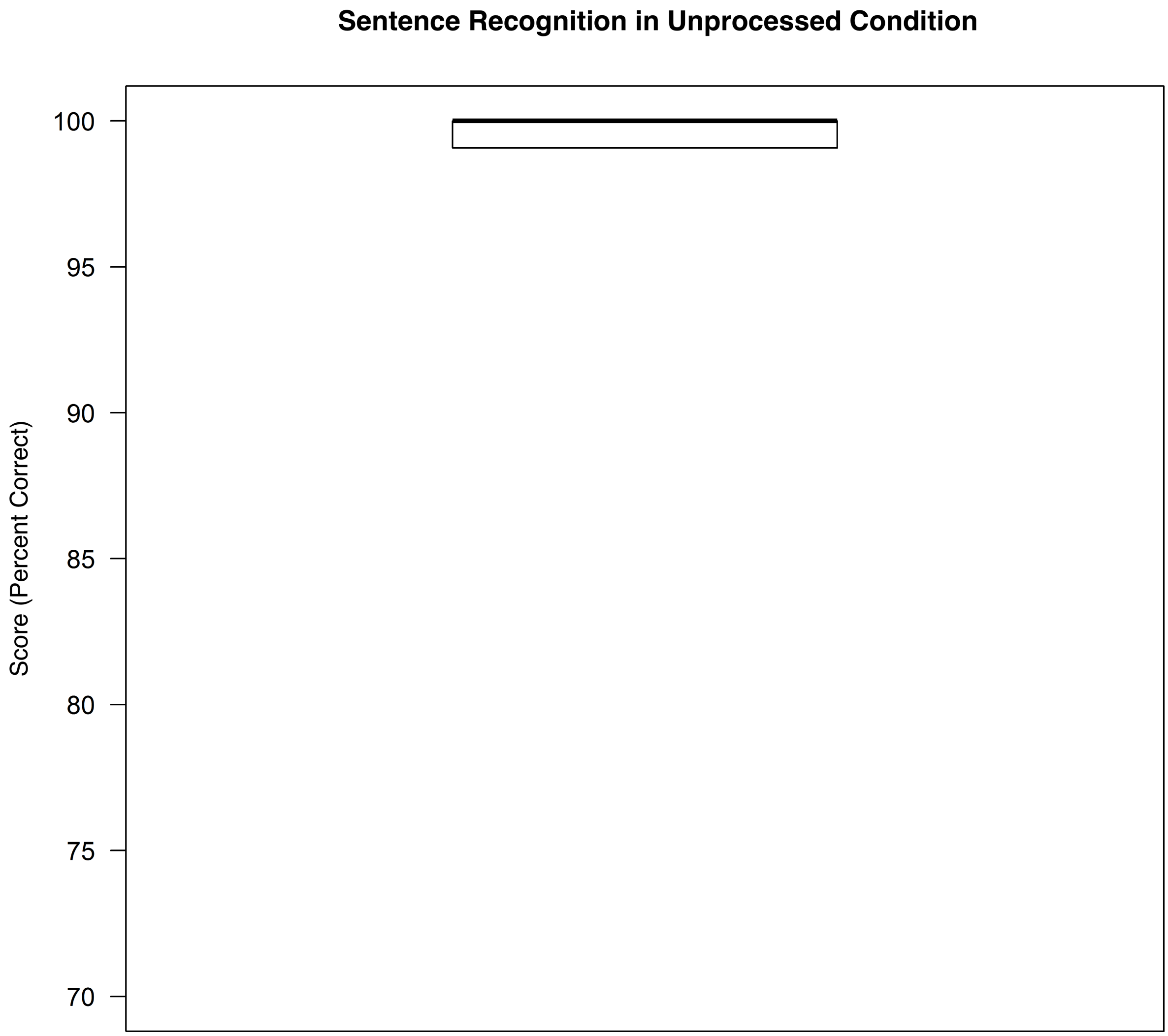

Figure 7 Box plot of sentence recognition scores (percent correct) for unprocessed stimuli, all talkers combined. The central line represents the median value and the box limits are the 25th and 75th percentiles.

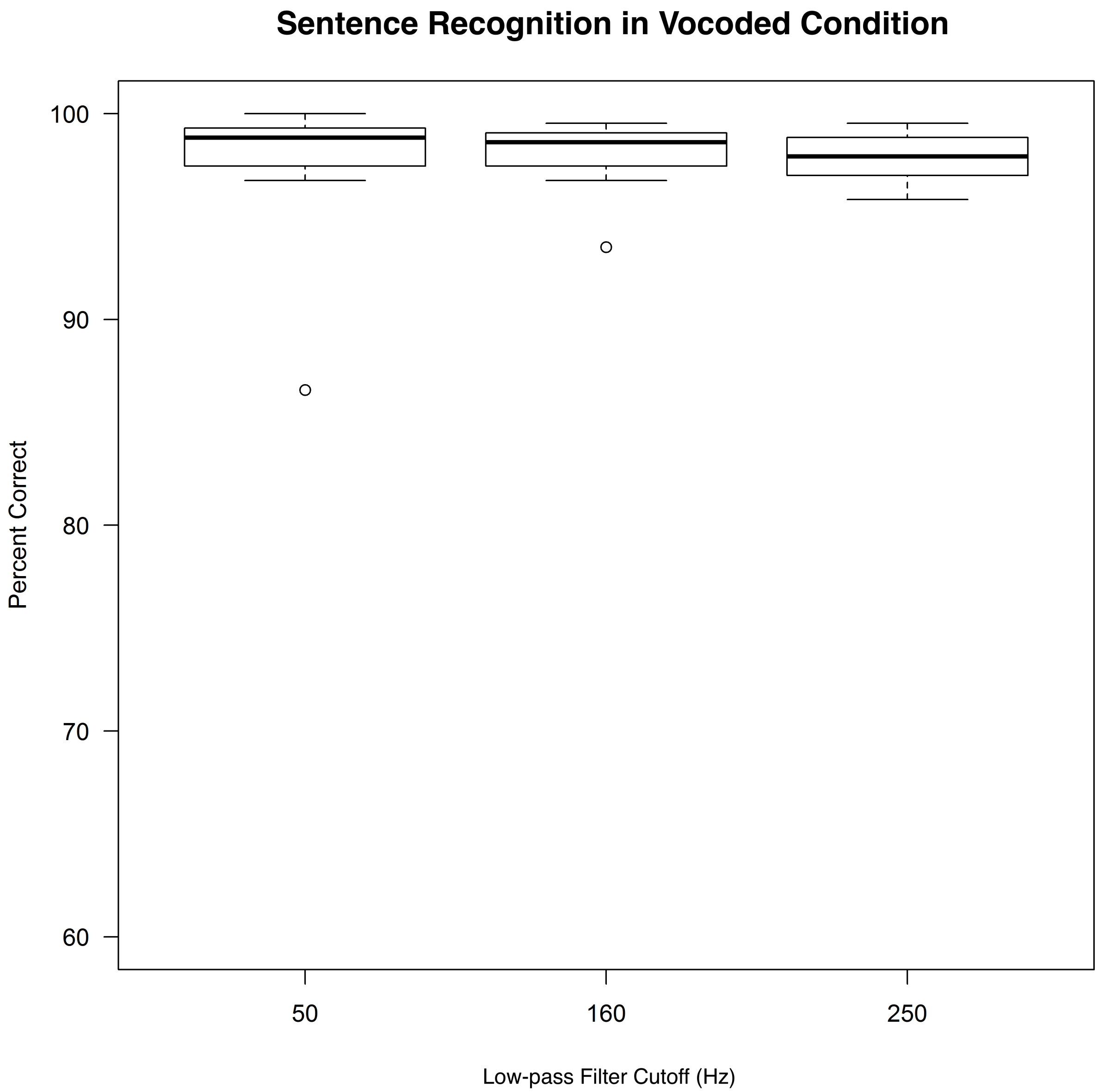

Figure 8 Box plot of sentence recognition scores (percent correct) for vocoded stimuli as a function of filter cutoff, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the lower 10th percentiles.

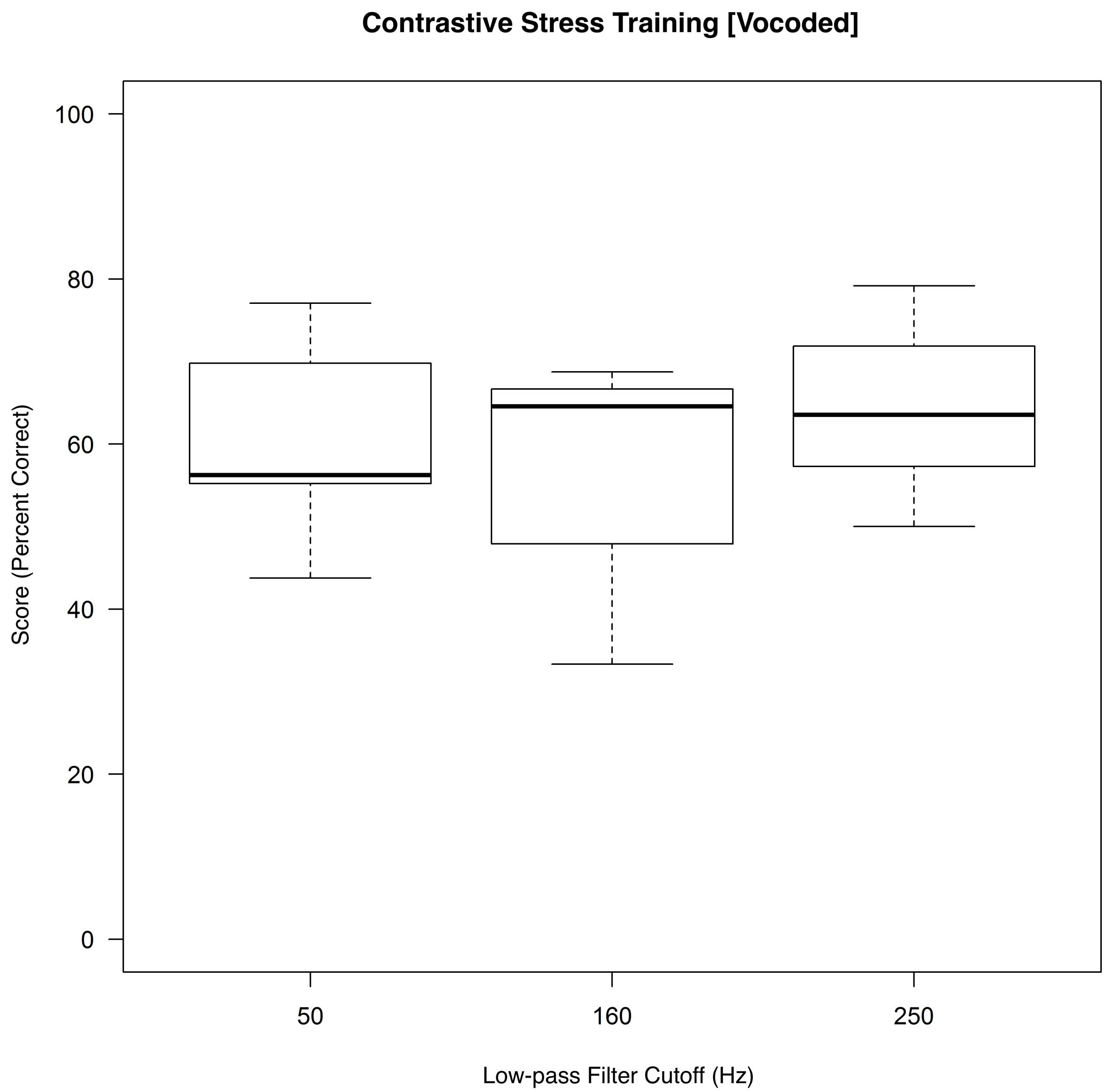

Figure 9 Box plot of contrastive stress training scores in the vocoded condition as a function of filter cutoff, all talkers combined. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles.

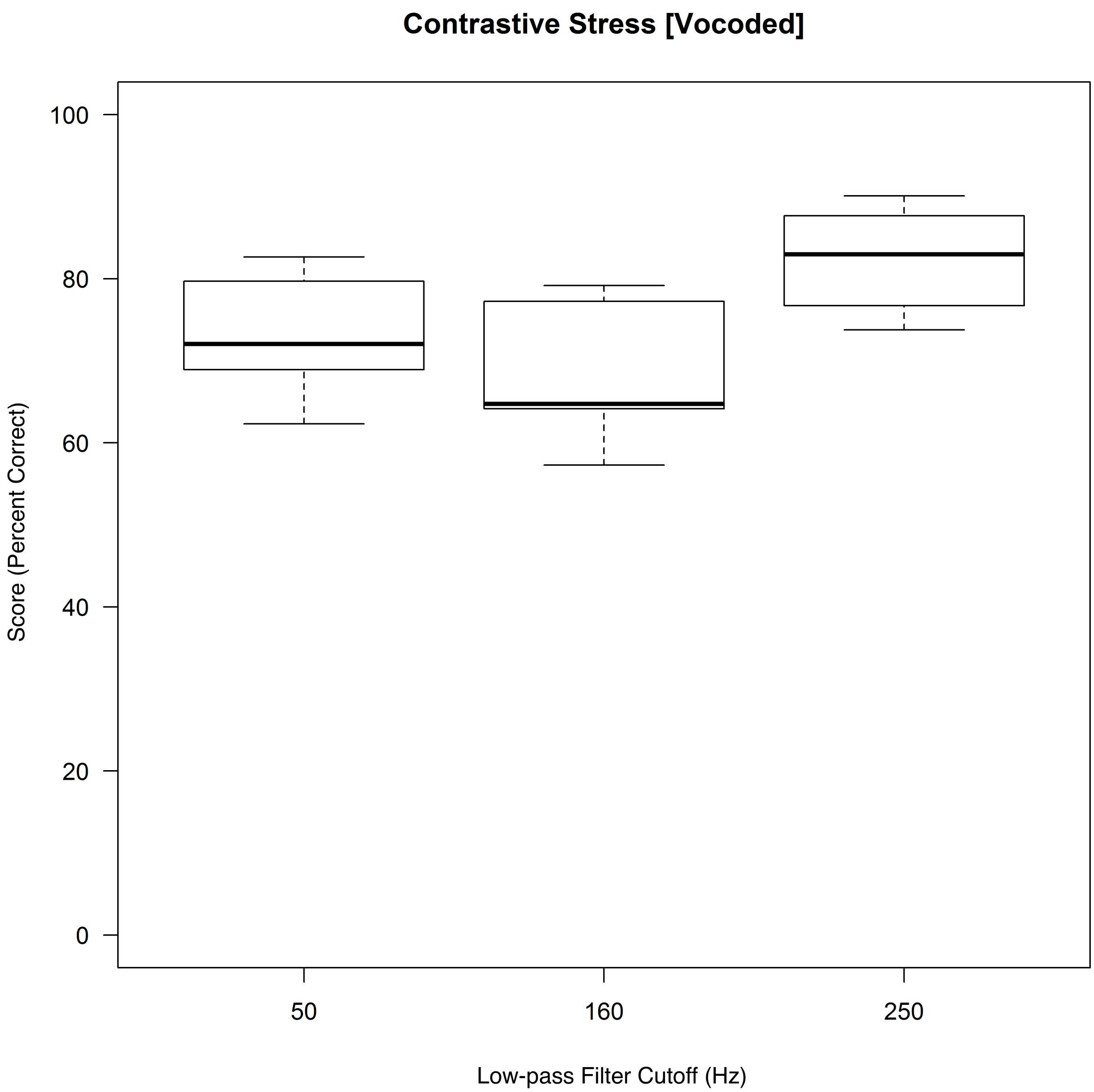

Figure 10 Box plot of contrastive stress test scores (percent correct) for vocoded stimuli as a function of filter cutoff. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles.

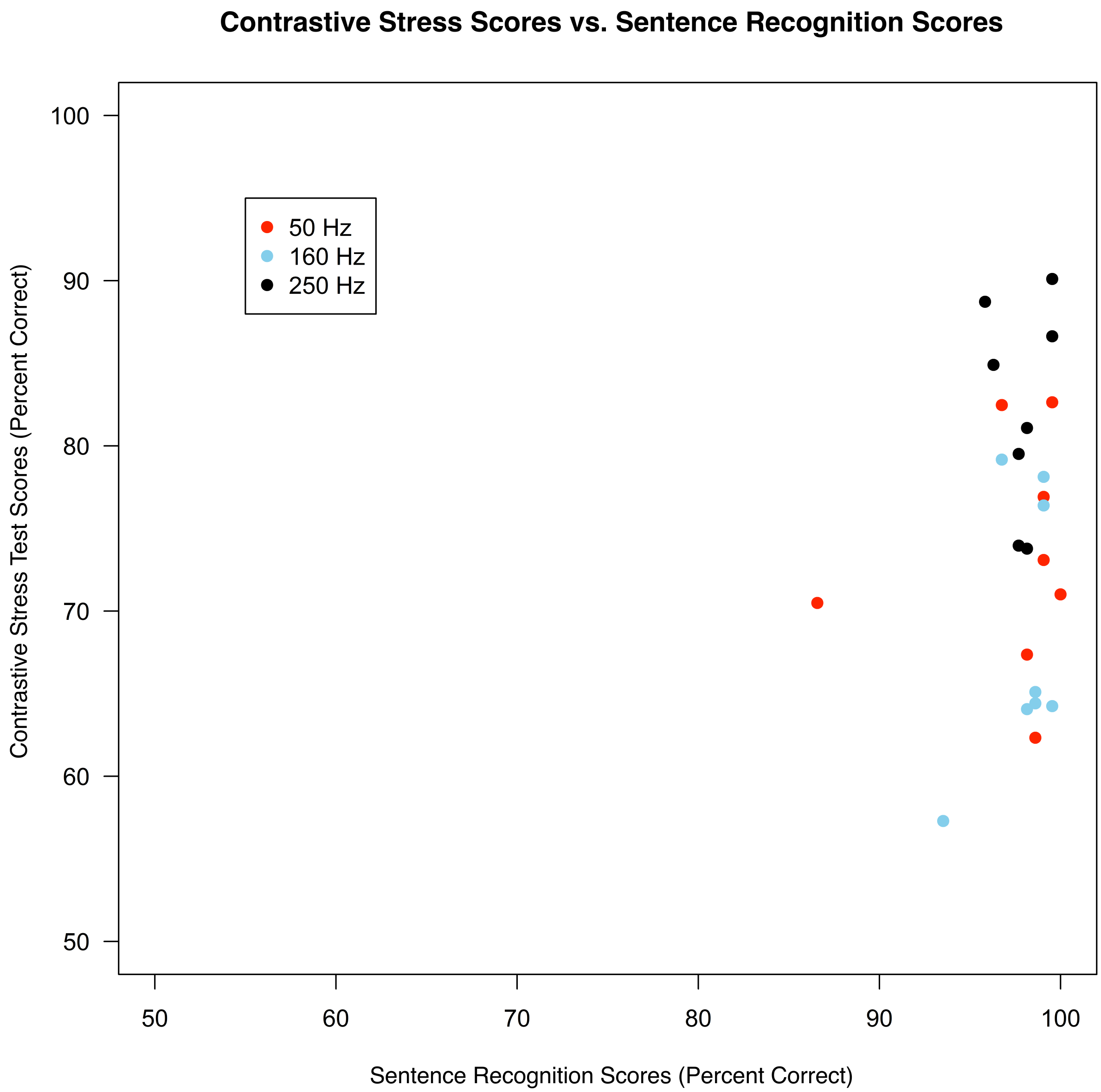

Figure 11 Scatter plot of contrastive stress test scores (percent correct) by sentence recognition scores (percent correct) for vocoded stimuli. The red circles represent the participant scores for the 50 Hz low-pass filter cutoff condition. The light blue circles represent the participant scores for the 160 Hz low-pass filter cutoff condition. The black circles represent the participant scores for the 250 Hz low-pass filter cutoff condition.

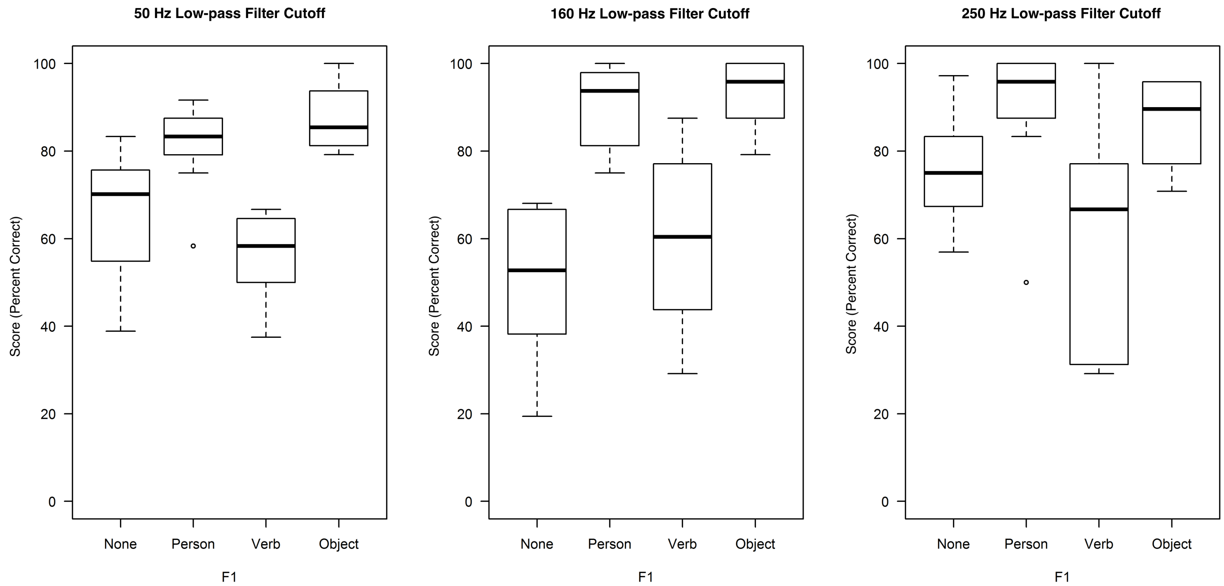

Figure 12 Box plot of contrastive stress test scores for each low-pass filter cutoff condition by each sentence stress condition, for talker F1. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the lower 10th percentiles.

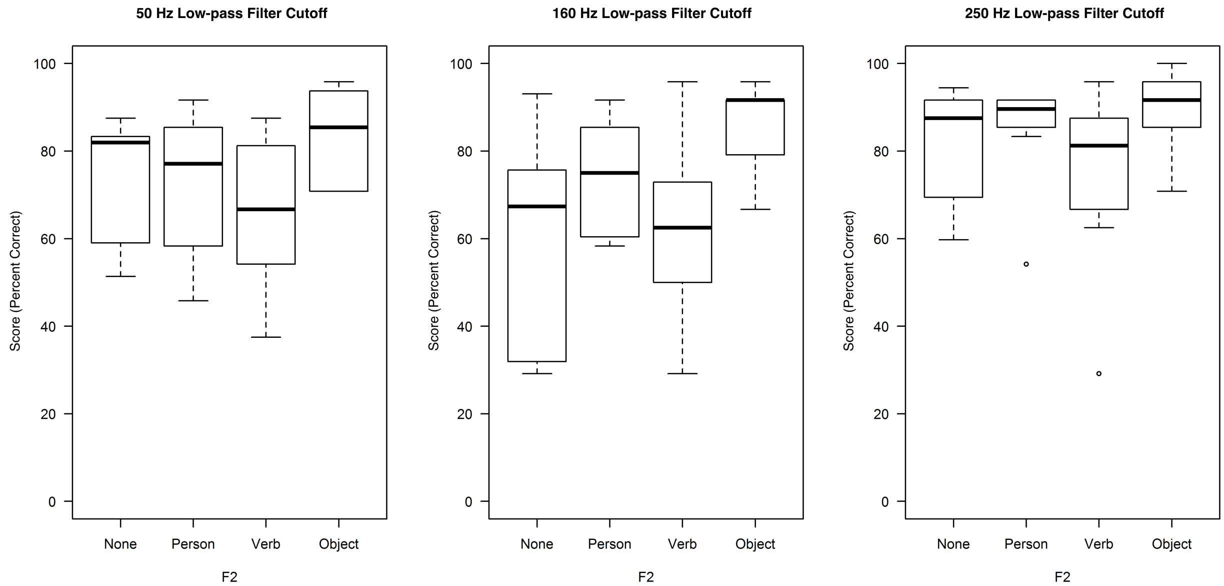

Figure 13 Box plot of contrastive stress test scores for each low-pass filter cutoff condition by each sentence stress condition, for talker F2. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the lower 10th percentiles.

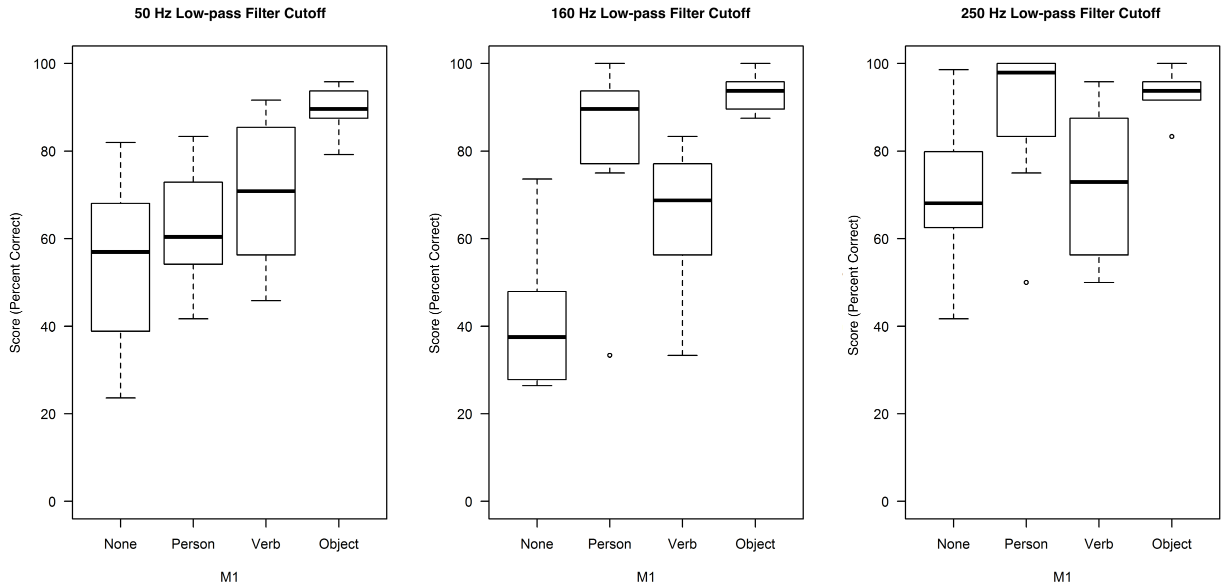

Figure 14 Box plot of contrastive stress test scores for each low-pass filter cutoff condition by each sentence stress condition, for talker M1. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the lower 10th percentiles.

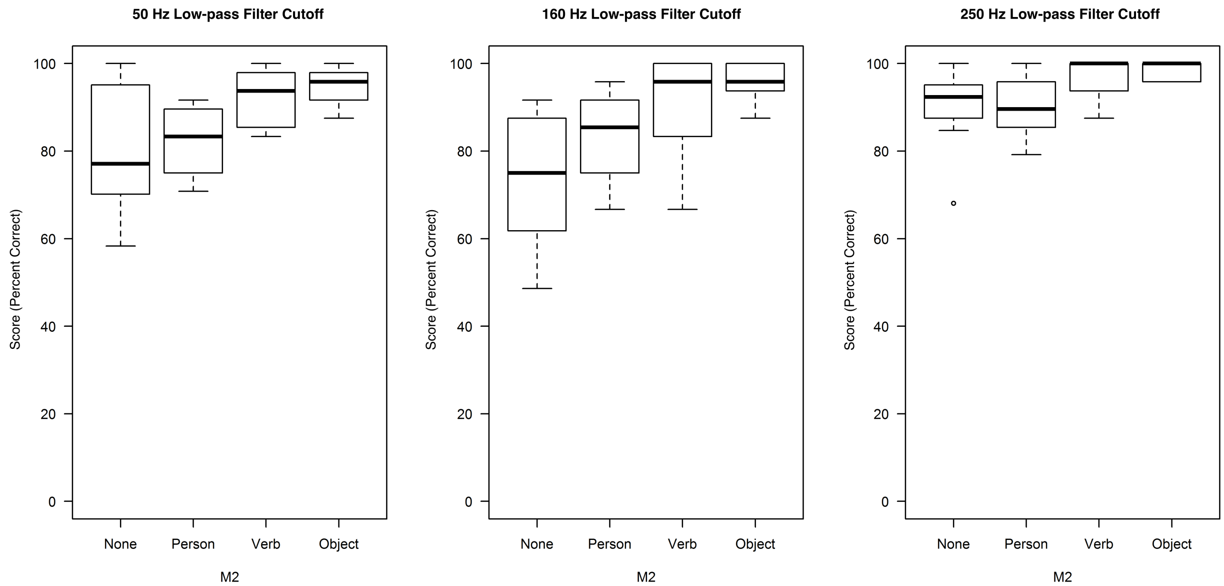

Figure 15 Box plot of contrastive stress test scores for each low-pass filter cutoff condition by each sentence stress condition, for talker M2. The central lines represent the median values and the box limits are the 25th and 75th percentiles. The lower fence is from the 10th to 25th percentiles and the upper fence is from the 75th to 90th percentiles. The circles represent outliers at the lower 10th percentiles.

VIII. Appendix

A. Sentence stimuli list, by sentence stress conditions.

Verb Stressed	**Person Stressed**	**Object Stressed**
Peg pushed the cart	Bob baked the bread	Bob poured the juice
Peg pulled the cart	Ted baked the bread	Bob poured the milk
Ted cut the bread	Ted drank the juice	Peg cleaned the desk
Ted ate the bread	Peg drank the juice	Peg cleaned the tub
Kate wrote the book	Peg cleaned the floor	Ted built the house
Kate read the book	Kate cleaned the floor	Ted built the bench
Peg poured the milk	Kate kicked the ball	Kate tied the shoes
Peg drank the milk	Bob kicked the ball	Kate tied the bows
Bob cleaned the floor	Bob read the book	Bob cut the cake
Bob swept the floor	Peg read the book	Bob cut the bread
Ted dried her hair	Ted called the dog	Kate pulled the rope
Ted combed her hair	Kate called the dog	Kate pulled the cart
Kate smelled the fruit	Peg parked the car	Peg painted the floor
Kate tasted the fruit	Bob parked the car	Peg painted the sun
Bob ran the race	Kate fed the dog	Ted cooked the meat
Bob walked the race	Ted fed the dog	Ted cooked the soup
Peg bought the food	Bob cooked the soup	Bob called the cat
Peg made the food	Kate cooked the soup	Bob called the dog
Ted kicked the ball	Ted pulled the cart	Ted turned the knob
Ted threw the ball	Bob pulled the cart	Ted turned the keys
Kate burned the cake	Peg painted the wall	Kate drove the cab
Kate baked the cake	Ted painted the wall	Kate drove the truck
Bob drove the car	Kate turned the knob	Peg bought the clock
Bob washed the car	Peg turned the knob	Peg bought the soap

B. Individual participant demographic information, gender and age, by participant code names for unprocessed speech group condition.

Code Name	Gender	Age (years)
NHU-01	F	19.08
NHU-02	F	21.27
NHU-03	F	22.53
NHU-04	F	32.41
NHU-05	F	35.11
NHU-06	M	20.91
NHU-07	F	24.19
NHU-08	F	18.07
NHU-09	M	26.98

* NHU: normal hearing participant, unprocessed condition group.
* F: Female
* M: Male

C. Individual participant demographic information, gender and age, by participant code names for processed speech group conditions.

Code Name	Low-pass Filter (Hz)	Gender	Age (years)
NHV-01	50	F	20.50
NHV-02	250	F	21.97
NHV-03	50	M	20.01
NHV-04	250	F	22.09
NHV-05	50	M	25.00
NHV-06	250	M	26.49
NHV-07	50	F	25.44
NHV-08	250	M	22.02
NHV-09	50	F	22.40
NHV-10	250	F	21.86
NHV-11	50	M	19.97
NHV-12	250	F	21.13
NHV-13	50	F	22.42
NHV-14	250	M	19.97
NHV-15	50	F	21.12
NHV-16	250	F	22.42
NHV-17	160	M	24.83
NHV-18	160	F	24.65
NHV-19	160	M	24.81
NHV-20	160	F	22.61
NHV-21	160	F	20.54
NHV-22	160	F	19.99

C. [Continued]

Code Name	Low-pass Filter (Hz)	Gender	Age
NHV-23	160	M	23.60
NHV-24	160	F	22.63

* NHV: normal hearing participant, processed condition group.
* F: Female
* M: Male

D. Individual participant condition assignments for sentence recognition and contrastive stress tests, by participant code names for unprocessed speech group condition.

	Sentence Recognition	Contrastive Stress Test			
	Corpus*	Talker 1	Talker 2	Talker 3	Talker 4
NHU-01	1	M1	M2	F1	F2
NHU-02	2	F2	F1	M2	M1
NHU-03	3	F1	M1	F2	M2
NHU-04	4	M2	F2	M1	F1
NHU-05	1	M1	M2	F1	F2
NHU-06	2	F2	F1	M2	M1
NHU-07	3	F1	M1	F2	M2
NHU-08	4	M2	F2	M1	F1
NHU-09	2	F2	F1	M2	M1

* Corpus: one of four compilations of 72 sentences

E. Individual participant condition assignments for sentence recognition and contrastive stress tests, by participant code names for processed speech group conditions.

	Word Recognition		Contrastive Stress Test			
	Low-pass Filter (Hz)	Corpus*	Talker 1	Talker 2	Talker 3	Talker 4
NHV-01	50	1	M1	M2	F1	F2
NHV-02	250	2	F2	F1	M2	M1
NHV-03	50	3	F1	M1	F2	M2
NHV-04	250	4	M2	F2	M1	F1
NHV-05	50	2	M1	M2	F1	F2
NHV-06	250	1	F2	F1	M2	M1
NHV-07	50	4	F1	M1	F2	M2
NHV-08	250	3	M2	F2	M1	F1
NHV-09	50	1	M1	M2	F1	F2
NHV-10	250	2	F2	F1	M2	M1
NHV-11	50	3	F1	M1	F2	M2
NHV-12	250	4	M2	F2	M1	F1
NHV-13	50	2	M1	M2	F1	F2
NHV-14	250	1	F2	F1	M2	M1
NHV-15	50	4	F1	M1	F2	M2
NHV-16	250	3	M2	F2	M1	F1
NHV-17	160	1	M1	M2	F1	F2
NHV-18	160	2	F2	F1	M2	M1
NHV-19	160	3	F1	M1	F2	M2
NHV-20	160	4	M2	F2	M1	F1
NHV-21	160	1	M1	M2	F1	F2

E. [Continued]

	Sentence Recognition		Contrastive Stress Test			
	Low-pass Filter (Hz)	Corpus*	Talker 1	Talker 2	Talker 3	Talker 4
NHV-22	160	2	F2	F1	M2	M1
NHV-23	160	3	F1	M1	F2	M2
NHV-24	160	4	M2	F2	M1	F1

* Corpus: one of four compilations of 72 sentences

F. Individual participant sentence recognition scores, unprocessed speech group condition.

	Sentence Recognition Score (Percent Correct)
NHU-01	99.08
NHU-02	99.08
NHU-03	100.00
NHU-04	100.00
NHU-05	100.00
NHU-06	100.00
NHU-07	100.00
NHU-08	99.08
NHU-09	100.00

G. Individual participant sentence recognition scores, processed speech groups condition.

	Low-pass Filter (Hz)	Sentence Recognition Score (Percent Correct)
NHV-01	50	99.53
NHV-02	250	96.75
NHV-03	50	99.53
NHV-04	250	99.08
NHV-05	50	96.30
NHV-06	250	99.53
NHV-07	50	98.13
NHV-08	250	100.00
NHV-09	50	97.68
NHV-10	250	99.08
NHV-11	50	97.68
NHV-12	250	98.60
NHV-13	50	98.13
NHV-14	250	98.13
NHV-15	50	95.83
NHV-16	250	99.05
NHV-17	160	96.73
NHV-18	160	99.05
NHV-19	160	98.13
NHV-20	160	98.60
NHV-21	160	98.58
NHV-22	160	93.50
NHV-23	160	99.53

G. [Continued]

	Low-pass Filter (Hz)	Sentence Recognition Score (Percent Correct)
NHV-24	160	98.13

H. Individual participant contrastive stress training scores, unprocessed speech group condition.

	Training Score (Percept Correct)
NHU-01	85.42
NHU-02	100.00
NHU-03	91.67
NHU-04	97.92
NHU-05	91.67
NHU-06	89.58
NHU-07	89.58
NHU-08	100.00
NHU-09	75.00

I. Individual participant contrastive stress training scores, processed speech group condition.

	Low-pass Filter (Hz)	Training Score (Percept Correct)
NHV-01	50	43.75
NHV-02	250	50.00
NHV-03	50	56.25
NHV-04	250	66.67
NHV-05	50	77.08
NHV-06	250	72.92
NHV-07	50	66.67
NHV-08	250	70.83
NHV-09	50	54.17
NHV-10	250	58.33
NHV-11	50	72.92
NHV-12	250	56.25
NHV-13	50	56.25
NHV-14	250	60.42
NHV-15	50	56.25
NHV-16	250	79.17
NHV-17	160	64.58
NHV-18	160	47.92
NHV-19	160	68.75
NHV-20	160	64.58
NHV-21	160	47.92
NHV-22	160	68.75
NHV-23	160	33.33

I. [Continued]

	Low-pass Filter (Hz)	Training Score (Percent Correct)
NHV-24	160	64.58

J. Individual participant contrastive stress experimental test scores for all talkers combined and separated, unprocessed speech group condition.

	Contrastive Stress Experimental Test Score (Percent Correct)				
	All Talkers	F1	F2	M1	M2
NHU-01	93.58	95.14	95.14	88.89	95.14
NHU-02	97.75	97.92	99.31	97.22	96.53
NHU-03	92.54	90.28	91.67	90.28	97.92
NHU-04	99.31	99.31	100.00	97.92	100.00
NHU-05	94.10	97.22	95.14	88.19	95.83
NHU-06	87.16	91.67	75.00	90.28	91.67
NHU-07	92.88	93.06	94.44	86.11	97.92
NHU-08	98.26	100.00	100.00	98.61	94.44
NHU-09	83.51	89.58	77.78	77.78	88.89

K. Contrastive stress experimental test score means by talker by place stress condition for unprocessed speech group condition.

		Contrastive Stress Experimental Test Score (Percent Correct)	
Talker	Place Stress Condition	Mean	Standard Deviation
F1	N	95.22	5.52
F2	N	95.22	5.43
M1	N	92.28	4.51
M2	N	99.23	1.41
F1	P	90.74	12.63
F2	P	80.09	26.00
M1	P	78.70	22.77
M2	P	86.11	12.15
F1	V	95.37	9.42
F2	V	93.98	16.55
M1	V	91.67	8.84
M2	V	93.52	6.63
F1	O	97.69	3.67
F2	O	92.59	8.78
M1	O	96.30	6.05
M2	O	94.91	6.17

* N: “no stress” place stress condition
* P: “person” place stress condition
* V: “verb” place stress condition
* O: “object” place stress condition

L. Individual participant contrastive stress experimental test scores for all talkers combined and separated, processed speech group condition.

		Contrastive Stress Experimental Test Score (Percent Correct)				
	Low-pass Filter (Hz)	All Talkers	F1	F2	M1	M2
NHV-01	50	70.49	73.61	81.94	53.47	72.92
NHV-02	250	90.11	90.97	84.03	90.28	95.14
NHV-03	50	82.47	76.39	80.56	79.86	93.06
NHV-04	250	86.63	81.25	90.28	78.47	96.53
NHV-05	50	76.91	68.75	75.69	72.92	90.28
NHV-06	250	84.90	80.56	85.42	78.47	95.14
NHV-07	50	82.64	72.22	87.50	74.31	96.53
NHV-08	250	81.08	68.75	88.19	75.69	91.67
NHV-09	50	71.01	75.69	64.58	62.50	81.25
NHV-10	250	79.52	71.53	84.03	67.36	95.14
NHV-11	50	73.09	73.61	72.92	64.58	81.25
NHV-12	250	73.96	71.53	68.75	74.31	81.25
NHV-13	50	62.33	60.42	62.50	47.22	79.17
NHV-14	250	73.79	64.58	66.67	70.83	93.06
NHV-15	50	67.36	61.11	63.89	58.33	86.11
NHV-16	250	88.72	88.19	90.28	84.72	91.67
NHV-17	160	78.13	75.69	86.11	65.28	85.42
NHV-18	160	79.17	75.69	66.67	81.25	93.06
NHV-19	160	76.39	68.75	75.00	68.06	93.75
NHV-20	160	64.06	68.06	59.03	56.94	72.22
NHV-21	160	65.10	64.58	68.75	48.61	78.47

L. [Continued]

	Low-pass Filter (Hz)	Contrastive Stress Experimental Test Score (Percent Correct)				
		All Talkers	F1	F2	M1	M2
NHV-22	160	64.41	55.56	77.08	54.86	70.14
NHV-23	160	57.29	58.33	43.06	50.69	77.08
NHV-24	160	64.24	59.03	56.25	57.64	84.03

M. Contrastive stress experimental test score means by talker by place stress condition for 50 Hz low-pass filter cutoff speech condition group.

		Contrastive Stress Experimental Test Score (Percent Correct)	
Talker	Place Stress Condition	Mean	Standard Deviation
F1	N	65.45	16.30
F2	N	73.44	14.59
M1	N	54.17	19.37
M2	N	80.38	14.87
F1	P	81.25	10.45
F2	P	72.40	16.81
M1	P	62.50	13.36
M2	P	82.29	7.95
F1	V	56.25	10.45
F2	V	66.15	17.74
M1	V	70.31	16.88
M2	V	92.19	6.84
F1	O	87.50	8.33
F2	O	83.33	11.14
M1	O	89.58	5.45
M2	O	94.79	4.31

* N: “no stress” place stress condition
* P: “person” place stress condition
* V: “verb” place stress condition
* O: “object” place stress condition

N. Contrastive stress experimental test score means by talker by place stress condition for 160 Hz low-pass filter cutoff speech condition group.

		Contrastive Stress Experimental Test Score (Percent Correct)	
Talker	Place Stress Condition	Mean	Standard Deviation
F1	N	50.35	17.86
F2	N	59.03	24.63
M1	N	40.80	15.85
M2	N	73.61	15.98
F1	P	90.10	9.69
F2	P	73.96	13.32
M1	P	81.77	21.24
M2	P	83.34	10.91
F1	V	59.90	21.35
F2	V	61.98	20.34
M1	V	65.10	16.21
M2	V	90.63	12.15
F1	O	93.23	8.01
F2	O	85.94	10.19
M1	O	93.23	4.42
M2	O	95.83	4.45

* N: “no stress” place stress condition
* P: “person” place stress condition
* V: “verb” place stress condition
* O: “object” place stress condition

O. Contrastive stress experimental test score means by talker by place stress condition for 250 Hz low-pass filter cutoff speech condition group.

		Contrastive Stress Experimental Test Score (Percent Correct)	
Talker	Place Stress Condition	Mean	Standard Deviation
F1	N	75.69	12.79
F2	N	81.42	13.71
M1	N	70.14	17.14
M2	N	89.76	9.82
F1	P	89.58	16.96
F2	P	84.90	12.78
M1	P	89.06	17.95
M2	P	90.10	7.02
F1	V	59.90	26.44
F2	V	74.48	21.18
M1	V	72.40	18.22
M2	V	96.88	4.85
F1	O	86.46	10.14
F2	O	89.58	9.45
M1	O	93.23	4.95
M2	O	98.44	2.16

* N: “no stress” place stress condition
* P: “person” place stress condition
* V: “verb” place stress condition
* O: “object” place stress condition

IX. Bibliography

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., & Clark, G. M. (1984). Speech processing studies using an acoustic model of a multiple‐channel cochlear implant. *The Journal of the Acoustical Society of America*, *76*(1), 104-110.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.28) [Computer program]. Available from http://www.praat.org/

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2014). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research, 322,* 151-162.

Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, *235*(1), 143-156.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142-2156.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America*, *102*(4), 2403-2411.

Dorman, M. F., & Wilson, B. S. (2004). The design and function of cochlear implants. *American Scientist*, *92*(5), 436-445.

Faulkner, A., Rosen, S., & Smith, C. (2000). Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, *108*(4), 1877-1887.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150-1163.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126-152.

Gonzalez, J., & Oliver, J. C. (2005). Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *The Journal of the Acoustical Society of America*, *118*(1), 461-470.

Hughes, M. (2013a). The basics of a cochlear implant. In *Objective measures in cochlear implants* (pp. 3-29). Plural Publishing.

Hughes, M. (2013b). Signal Delivery. In *Objective measures in cochlear implants* (pp. 31-41). Plural Publishing.

Laneau, J., Moonen, M., & Wouters, J. (2006). Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *The Journal of the Acoustical Society of America*, *119*(1), 491-506.

Loebach, J. L. (2007). *Cochlear implant simulations: a tutorial on generating acoustic simulations for research* (No. 28, pp. 359-368). Progress report.

McRoberts, G. W., Studdert-Kennedy, M., & Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Attention, Perception, & Psychophysics*, *57*(2), 159-174.

Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M. (2011). The perception of sentence stress in cochlear implant recipients. *Ear & Hearing*, 32(4), 459-467.

Meister, H., Landwehr, M., Pyschny, V., Walger, M. & von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology, 48*(1), 38-48.

Most, T. & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. *Journal of Deaf Studies and Deaf Education, 12*(3), 350-361.

Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends in Hearing, 18,* 2331216514553783.

Peng, S.C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification, 16*(2), 67-82.

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing*, *26*(5), 451-460.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royals Society of London: Biological Sciences*, *336*(1278), 367-373.

Schvartz, K. C. & Chatterjee, M. (2012). Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech. *Ear & Hearing, 33*(3), 411-420.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303.

Souza, P., & Rosen, S. (2009). Effects of envelope bandwidth on the intelligibility of sine-and noise-vocoded speech. *The Journal of the Acoustical Society of America*, *126*(2), 792-805.

Watson, P. J., & Schlauch, R. S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *American Journal of Speech-Language Pathology*, *17*(4), 348-355.

Whitmal III, N. A., Poissant, S. F., Freyman, R. L., & Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *The Journal of the Acoustical Society of America*, *122*(4), 2376-2388.

Wilson, B. S. (2015). Getting a decent (but sparse) signal to the brain for users of cochlear implants. *Hearing Research*, *322*, 24-38.

Wilson, B. S., Dorman, M. F., Gifford, R. H., & McAlpine, D. (2016). Cochlear implant design considerations. In *Pediatric cochlear implantation* (pp. 3-23). Springer New York.

Xu, L., & Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hearing Research*, *242*(1), 132-140.